

A/Box
seg

ARNOLD, WHITE & DURKEE

A PROFESSIONAL CORPORATION

Attorneys at Law

POST OFFICE BOX 4433

HOUSTON, TX 77210-4433

750 BERING DRIVE, SUITE 400

HOUSTON, TX 77057-2198

TELEPHONE (713) 787-1400

FACSIMILE (713) 789-2679

February 19, 1998

800 QUAKER TOWER
321 NORTH CLARK STREET
CHICAGO, IL 60610-4714
TELEPHONE (312) 744-0090
FACSIMILE (312) 755-4489

2001 JEFFERSON DAVIS HIGHWAY
SUITE 401
ARLINGTON, VA 22202-3604
TELEPHONE (703) 415-1720
FACSIMILE (703) 415-1728

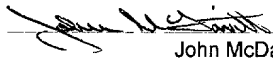
FILE: UTXC:506

1900 ONE AMERICAN CENTER
600 CONGRESS AVENUE
AUSTIN, TX 78701-3248
TELEPHONE (512) 418-3000
FACSIMILE (512) 474-7577

155 LINFIELD DRIVE
MENLO PARK, CA 94025-3741
TELEPHONE (415) 614-4500
FACSIMILE (415) 614-4599

4850 FIRST BANK PLACE
601 SECOND AVE SOUTH
MINNEAPOLIS, MN 55402-4320
TELEPHONE (612) 321-2800
FACSIMILE (612) 321-9600

WRITER'S DIRECT DIAL:
(512) 418-5616

EXPRESS MAIL MAILING LABEL
NUMBER EM 423 823 788 US
DATE OF DEPOSIT February 19, 1998
I hereby certify that this paper or fee is being 20, 1998 deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington D C. 20231.
 John McDavitt

BOX PATENT APPLICATION

Assistant Commissioner for Patents
Washington, DC 20231

RE: *U. S. Patent Application Entitled: MODIFIED RETINOBLASTOMA TUMOR SUPPRESSOR PROTEINS - Hong-Ji Xu et al. (UTMDACC:506)*

Sir:

Transmitted herewith for filing is a 258 page patent specification including 43 claims and an abstract. Also included are Formal Drawings for Figures 1-5, which represent 9 drawings on 5 sheets. The specification and drawings constitute the application of Hong-Ji Xu, Shi-Xue Hu, William F. Benedict and Yunli Zhou for the captioned application.

Also transmitted herewith is a diskette containing the computer-readable form of those sequences in the specification, a Statement as Required Under 37 C.F.R. § 1.821(f), and a separate paper copy of the sequence listing.

Please note that this application is filed without an inventor Declaration and Assignment, a Declaration Claiming Small Entity Status, a Power of Attorney, and filing fees. Pursuant to 37 C.F.R. § 1.53(b) and (d), the Applicant requests the Patent and Trademark Office to accept this application and accord a serial number and filing date as of the date this application is

ARNOLD, WHITE & DURKEE

Assistant Commissioner for Patents

February 19, 1998

Page 2

deposited with the U.S. Postal Service for Express Mail. Further, the Applicant requests that the NOTICE OF MISSING PARTS-FILING DATE GRANTED pursuant to 37 C.F.R. § 1.53(d) be sent to the undersigned Applicant's representative.

Please date stamp and return the enclosed postcard to evidence receipt of this application.

Respectfully submitted,

David W. Hibler

David W. Hibler

Reg. No. 41,071

Agent for Applicants


DWH:jmg

Encl:

SECRET

PATENT
UTXC:506

APPLICATION FOR UNITED STATES LETTERS PATENT
for
MODIFIED RETINOBLASTOMA TUMOR SUPPRESSOR PROTEINS
by
Hong-Ji Xu
Shi-Xue Hu
William F. Benedict
and
Yunli Zhou

EXPRESS MAIL MAILING LABEL	
NUMBER	EM 423 823 788 US
DATE OF DEPOSIT	February 19, 1998
I hereby certify that this paper or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington D.C. 20231.	
	
John McDavitt	

BACKGROUND OF THE INVENTION

The present application claims the priority of co-pending U.S. Provisional Patent Application Serial No. 60/038,118, filed February 20, 1997, incorporated herein by reference in its entirety without disclaimer. The government owns rights in the present invention pursuant to grant numbers R01-CA 67274 and R01-EY 06195 from the National Institutes of Health, and grant number ATP004949018 from the Texas Higher Education Coordinating Board.

1. Field of the Invention

The present invention relates generally to the field of molecular and cellular biology. More particularly, it concerns modifications of the retinoblastoma tumor suppressor. The present invention further relates to the use of the instant modified retinoblastoma tumor suppressors in situations where providing a tumor suppressor or normal cell growth suppressor is indicated.

2. Description of Related Art

Cancers and tumors are the second most prevalent cause of death in the United States, causing approximately 450,000 deaths per year. One in three Americans will develop cancer, and one in five will die of cancer (Scientific American Medicine, part 12, I, 1, section dated 1987). While substantial progress has been made in identifying some of the likely environmental and hereditary causes of cancer, the statistics for the cancer death rate indicates a need for substantial improvement in the therapy for cancer and related diseases and disorders.

A number of genes have been implicated in the etiology of cancer. These genes have been identified in connection with hereditary forms of cancer, and in a large number of well-studied tumor cells. Study of cancer genes has helped provide some understanding of the process of tumorigenesis. While a great deal more remains to be learned about cancer genes, the presently known cancer genes serve as useful models for understanding tumorigenesis. Cancer genes are broadly classified into "oncogenes" which, when activated, promote tumorigenesis, and "tumor suppressor genes" which, when damaged, fail to suppress tumorigenesis. While these classifications provide a useful method for conceptualizing tumorigenesis, it is also possible that

a particular gene may play differing roles depending upon the particular allelic form of that gene, its regulatory elements, the genetic background and the tissue environment in which it is operating.

5 The oncogenes are somatic cell genes that are mutated from their wild-type alleles (the art refers to these wild-type alleles as protooncogenes) into forms which are able to induce tumorigenesis under certain conditions. There is presently a substantial literature on known and putative oncogenes and the various alleles of these oncogenes. For example, the oncogenes *ras* and *myc* are considered as models for understanding oncogenic processes in general. The *ras* oncogene is believed to encode a cytoplasmic protein, and the *myc* oncogene is believed to encode a nuclear protein. Neither the *ras* oncogene nor the *myc* oncogene alone is able to induce full transformation of a normal cell into a tumor cell, but full tumorigenesis usually occurs when both the *ras* and *myc* oncogenes are present and expressed together in the same cell (Weinberg, 1989). Such collaborative effects have been observed between a number of other studied oncogenes.

15 The collaborative model of oncogene tumorigenesis must be qualified by the observation that a cell expressing the *ras* oncogene that is surrounded by normal cells does not undergo full transformation. However, if most of the surrounding cells are also *ras*-expressing, then the *ras* oncogene alone is sufficient to induce tumorigenesis in a *ras*-expressing cell. This observation validates the multiple hit theory of tumorigenesis because a change in the tissue environment of the cell hosting the oncogene may be considered a second hit. An alternative and equally valid hypothesis is that events that collaborate with the activation of an oncogene such as *ras* or *myc* may include the inactivation of a negative regulatory factor or factors, *i.e.*, a tumor suppressor protein (Weinberg, 1989; Goodrich *et al.*, 1992a).

25 Tumor suppressor genes are genes that, in their wild-type alleles, express proteins that suppress abnormal cellular proliferation. When the gene coding for a tumor suppressor protein is mutated or deleted, the resulting mutant protein or the complete lack of a tumor suppressor protein may fail to correctly regulate cellular proliferation. This can lead to abnormal cellular

proliferation, particularly if there is already existing damage to the cellular regulatory mechanism. The lack of control of cellular proliferation has been linked to the development of a wide variety of human cancers (Weinberg, 1991). A number of well-studied human tumors and tumor cell lines have been shown to have missing or nonfunctional tumor suppressor genes.

5
Examples of tumor suppressor genes and candidate tumor suppressor genes include, but are not limited to, the retinoblastoma (RB) gene (Friend *et al.*, 1986; Fung *et al.*, 1987; Lee *et al.*, 1987a), the wild-type p53 gene (Finlay *et al.*, 1989; Baker *et al.*, 1990), the deleted in colon carcinoma (DCC) gene (Fearon *et al.*, 1990a; 1990b), the neurofibromatosis type 1 (NF-1) gene
10 (Wallace *et al.*, 1990; Viskochil *et al.*, 1990; Cawthon *et al.*, 1990), the Wilms tumor (WT-1) gene (Call *et al.*, 1990; Gessler *et al.*, 1990; Pritchard-Jones *et al.*, 1990), the von Hippel-Lindau (VHL) disease tumor suppressor gene (Duan *et al.*, 1995), the Maspin (Zou *et al.*, 1994), Brush-1 (Schott *et al.*, 1994) and BRCA 1 genes (Miki *et al.*, 1994; Futreal *et al.*, 1994) for breast cancer, and the multiple tumor suppressor (MTS) or p16 gene (Serrano *et al.*, 1993; Kamb *et al.*, 1994).
15 The list of putative tumor suppressor genes is large and growing, with the total number of tumor suppressor genes expected to be well beyond 50 (Knudson, 1993).

The first tumor suppressor gene identified was the retinoblastoma (RB) gene, which causes the hereditary retinoblastoma (Knudson, 1971; Murphree and Benedict, 1984; Knudson,
20 1985). The retinoblastoma (RB) gene, which was cloned in the middle 1980s, is one of the best studied tumor suppressor genes. The size of the RB gene complementary DNA (cDNA), about 4.7 kb, permits ready manipulation of the gene, and has led to the insertion of the RB gene into a number of cell lines. The RB gene has been shown to be missing or defective in a majority of retinoblastomas, sarcomas of the soft tissues and bones, and in approximately 20 to 40 percent of
25 breast, lung, prostate and bladder carcinomas (Lee *et al.*, WO 90/05180; Bookstein *et al.*, 1991; Benedict *et al.*, 1990).

The most direct proof that the cloned RB gene is indeed a tumor suppressor gene is the observed recovery of tumor suppression function in RB-minus tumor cells from the introduction
30 of a cloned intact copy of the RB gene. A number of reports have indicated that replacement of

the normal RB gene in RB-defective tumor cells from disparate types of human cancers could suppress their tumorigenic activity in nude mice (Huang *et al.*, 1988; Goodrich and Lee, 1993; Zhou *et al.*, 1994b). The tumor cell lines studied were derived from widely disparate types of human cancers such as the retinoblastoma, osteosarcoma, carcinomas of the bladder, prostate, breast and lung.

While it was observed that introduction of a functional wild-type, full-length retinoblastoma gene (RB¹¹⁰) into an RB-minus tumor cell "normalizes" the cell, it was not expected that tumor cells which already have normal RB¹¹⁰ gene expression ("RB⁺") would respond to RB¹¹⁰ gene therapy, because it was presumed that adding additional RB expression could not correct a non-RB genetic defect. This has in fact been shown for the case of the RB⁺ osteosarcoma cell line U-2 OS, where the introduction of an extra p110^{RB} coding gene did not change the neoplastic phenotype (Huang *et al.*, 1988). Thus, there remains a need for a broad-spectrum tumor suppressor gene for treating abnormally proliferating cells having any type of genetic defect.

The RB¹¹⁰ cDNA open reading frame sequence (McGee *et al.*, 1989) contains a second in-frame AUG codon located in exon 3, at nucleotides 355-357. The protein initiated from this second AUG codon lacks the N-terminal 112 amino acid residues of the full-length RB protein, and is termed pRB⁹⁴ (Xu *et al.*, 1994b). In U.S. Patent 5,496,731 (incorporated herein by reference), the inventors showed that RB-defective tumor cells expressing exogenous pRB⁹⁴ did not progress through the cell cycle, as evidenced by their failure to incorporate [³H]-thymidine into DNA. In contrast, the percent of tumor cells undergoing DNA replication were only slightly lower in cells producing the exogenous pRB¹¹⁰ (the wild-type pRB protein) than in cells that were RB⁻. Even more striking was that the pRB⁹⁴ expression also significantly reduced colony formation of two RB⁺ (with normal RB alleles) tumor cell lines examined, namely the fibrosarcoma cell line, HT1080, and the cervical carcinoma cell line, HeLa (Xu *et al.*, 1994b), while no such effects were observed when an additional pRB¹¹⁰-coding gene(s) was introduced

by transfection using plasmid vectors (Fung *et al.*, 1993) or by microcell fusion (Anderson *et al.*, 1994).

However, there is a paucity of tumor suppressor proteins in the art which have all of the properties necessary to facilitate their use in the treatment of diseases, particularly cancer.

SUMMARY OF THE INVENTION

The modified retinoblastoma tumor suppressors of the present invention overcome the shortcomings of those described in the art, providing a broad spectrum tumor suppressor with surprising beneficial effects.

The present invention provides broad-spectrum modified retinoblastoma tumor suppressor proteins that are surprisingly at least as effective, and in most cases more effective, than the corresponding wild-type retinoblastoma tumor suppressor proteins in inhibiting cell growth. In particular embodiments, the invention provides retinoblastoma tumor suppressor proteins that have a modified N-terminal region. The invention further provides methods of making and using the modified retinoblastoma tumor suppressor proteins, particularly in circumstances wherein cell growth inhibition is desired. Thus the present invention provides methods for treating diseases, as exemplified by, but not limited to cancer, that are characterized by abnormal cellular proliferation.

A broad-spectrum tumor suppressor gene is a genetic sequence coding for a protein that, when inserted into and expressed in an abnormally proliferating host cell, *e.g.*, a tumor cell, suppresses abnormal proliferation of that cell irrespective of the cause of the abnormal proliferation.

Thus, the invention provides an isolated DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴ or pRB⁵⁶, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification. The

terms "pRB⁹⁴" and "pRB⁵⁶" refer to retinoblastoma proteins that have a molecular weight of 94 kDa and 56 kDa, respectively. As understood in the art, the pRB⁹⁴ and pRB⁵⁶ retinoblastoma proteins are fragments of the full length wild-type retinoblastoma protein that have 112 and 379 contiguous amino acids deleted from the N-terminus, respectively.

5

The term "N-terminal", or "N-terminal region", as used herein, will be understood to refer to the region of a protein corresponding to as much as the first approximately 40% of the amino acid sequence. Thus, these terms will be understood to include up to about the first 5%, the first 10%, the first 15%, the first 20%, the first 25%, the first 30% or the first 35% of the amino acid sequence of a protein. However, these values are only approximations, and therefore will be understood to include intermediate values, such as 2%, 3%, 6%, 7%, 11%, 13%, 17%, 18%, 22%, 26%, 33%, 37%, 38%, 41%, 42% and the like.

10

The term "modified", as used herein, refers to deletions and/or mutations of the wild-type protein sequence. In certain embodiments, it may also refer to insertion of a heterologous amino acid or amino acids into the wild-type protein sequence. In yet other aspects, the term may refer to post-translational alteration of the wild-type amino acid sequence.

15

In a further embodiment of the invention, the gene encodes a modified retinoblastoma tumor suppressor protein comprising an N-terminal region that comprises a first sequence region from which at least one amino acid has been deleted. The deletion may produce a modified retinoblastoma tumor suppressor protein with a biological activity equal to, or in certain embodiments, greater than the biological activity of the corresponding wild-type retinoblastoma tumor suppressor protein.

20

25

In a particular embodiment of the invention the gene encodes a modified retinoblastoma tumor suppressor protein wherein at least two amino acids have been deleted from the first sequence region. In other embodiments of the invention at least about five amino acids, at least about ten amino acids, at least about 25 amino acids, at least about 50 amino acids, at least about 75 amino acids or at least about 100 amino acids have been deleted from the first sequence

30

region. It will be understood that intermediate deletion sizes are also contemplated, such as, but not limited to, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99 amino acids and the like.

In other aspects of the invention, the gene encodes a modified retinoblastoma tumor suppressor protein wherein at least about 150 amino acids, at least about 200 amino acids, at least about 250 amino acids, at least about 300 amino acids or at least about 370 amino acids have been deleted from the first sequence region. However, intermediate sized deletions are also provided, exemplified by, but not limited to, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 371, 372, 373, 374, 375, 376, 377 or 378 amino acid deletions. Other intermediate values are disclosed throughout the specification.

In one embodiment of the invention the gene encodes a modified retinoblastoma tumor suppressor protein comprising an N-terminal region that comprises at least a first sequence region located between about amino acid 1 and about amino acid 50 from which at least one

amino acid has been deleted. It will be understood that "between about amino acid 1 and about amino acid 50" includes amino acid 1 and amino acid 50, and it is thus so with other deletions described herein. Amino acid 1 is the N-terminal amino acid, and the numbers increase toward the C-terminus.

5

In further embodiments of the invention, the first sequence region is located between about amino acid 51 and about amino acid 100, between about amino acid 101 and about amino acid 150, between about amino acid 151 and about amino acid 200, between about amino acid 201 and about amino acid 250 or between about amino acid 251 and about amino acid 300.

10

In other embodiments of the present invention, the gene encodes a modified retinoblastoma tumor suppressor protein wherein the first sequence region is located between about amino acid 1 and about amino acid 100, between about amino acid 51 and about amino acid 150, between about amino acid 101 and about amino acid 200, between about amino acid 151 and about amino acid 250 or between about amino acid 201 and about amino acid 300.

In a particular aspect of the invention the gene encodes a modified retinoblastoma tumor suppressor protein wherein the first sequence region is located between about amino acid 1 and about amino acid 150. In additional aspects of the invention the first sequence region is located between about amino acid 51 and about amino acid 200, between about amino acid 101 and about amino acid 250 or between about amino acid 151 and about amino acid 300.

In further embodiments of the invention the gene encodes a modified retinoblastoma tumor suppressor protein wherein the first sequence region is located between about amino acid 1 and about amino acid 200, between about amino acid 51 and about amino acid 250, between about amino acid 101 and about amino acid 300, between about amino acid 1 and about amino acid 250, between about amino acid 51 and about amino acid 300, between about amino acid 1 and about amino acid 300 or between about amino acid 1 and about amino acid 370.

In yet another aspect of the invention the modified retinoblastoma tumor suppressor protein is a modified retinoblastoma protein wherein about amino acid 2 through about amino acid 34 have been deleted from the first sequence region. The location of these particular amino acids is in reference to the human wild-type retinoblastoma protein, but will be understood to correspond to analogous regions of homologous retinoblastoma proteins. In yet another aspect of the invention about amino acid 2 through about amino acid 55 have been deleted from the first sequence region. In still another aspect of the invention about amino acid 2 through about amino acid 78 have been deleted from the first sequence region. In a particular aspect of the invention about amino acid 2 through about amino acid 97 have been deleted from the first sequence region. In an additional aspect of the invention about amino acid 2 through about amino acid 148 have been deleted from the first sequence region.

In another embodiment of the invention the modified retinoblastoma tumor suppressor protein is a modified retinoblastoma protein wherein about amino acid 31 through about amino acid 107 have been deleted from the first sequence region. In another embodiment of the invention about amino acid 77 through about amino acid 107 have been deleted from the first sequence region. In a further embodiment of the invention about amino acid 111 through about amino acid 181 have been deleted from the first sequence region. In yet another embodiment of the invention about amino acid 111 through about amino acid 241 have been deleted from the first sequence region. In still another embodiment of the invention about amino acid 181 through about amino acid 241 have been deleted from the first sequence region. In a particular embodiment of the invention about amino acid 242 through about amino acid 300 have been deleted from the first sequence region.

In one aspect of the invention the N-terminal region of the modified retinoblastoma tumor suppressor protein further comprises at least a second sequence region from which at least one amino acid has been deleted. In a particular aspect of the invention, about amino acid 2 through about amino acid 34, and about amino acid 76 through about amino acid 112 have been deleted. In a further aspect of the invention about amino acid 2 through about amino acid 55, and about amino acid 76 through about amino acid 112 have been deleted.

Another embodiment of the invention provides a DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification wherein the gene encodes a modified retinoblastoma tumor suppressor protein comprising at least a first N-terminal mutation, and wherein the modified retinoblastoma tumor suppressor protein has an increased biological activity in comparison to the biological activity of the corresponding wild type retinoblastoma tumor suppressor protein. In one embodiment of the invention the gene encodes a modified retinoblastoma protein comprising a mutation at position 111. In another embodiment of the invention the modified retinoblastoma protein comprises glycine at position 111 in place of aspartic acid.

In a further embodiment of the invention the modified retinoblastoma tumor suppressor protein comprises at least a second N-terminal mutation. In yet another embodiment of the invention the gene encodes a modified retinoblastoma protein comprising a mutation at position 111 and a mutation at position 112. In still another embodiment of the invention the modified retinoblastoma protein comprises glycine at position 111 in place of aspartic acid, and aspartic acid at position 112 in place of glutamic acid. In a particular embodiment of the invention the gene encodes a modified retinoblastoma tumor suppressor protein comprising an N-terminal region from which at least one amino acid has been deleted, and which contains at least one amino acid mutation.

In one aspect of the invention the gene encodes a modified retinoblastoma tumor suppressor protein that comprises a contiguous amino acid sequence from at least about position 370 to about position 928 of SEQ ID NO:2. In another aspect of the invention the gene encodes a modified retinoblastoma tumor suppressor protein that comprises a contiguous amino acid sequence from at least about position 3 to about position 928 of SEQ ID NO:2. When used in this context, "a contiguous amino acid sequence" will be understood to be a contiguous amino acid sequence of at least about 8, about 10, about 12, about 15, about 20, about 25, about 50 or about 100 amino acids and so on up to the full length amino acid sequence.

In a further aspect of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:29. In yet another aspect of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2691 of SEQ ID NO:28. When used herein in this context, "a contiguous nucleic acid sequence" will be understood to be a contiguous nucleic acid sequence of at least about 8, about 10, about 12, about 15, about 17, about 20, about 25, about 50 or about 100 nucleotides and so on up to the full length nucleotide sequence.

In still another aspect of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:31. In a particular aspect of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2628 of SEQ ID NO:30. In an additional aspect of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:33.

In another embodiment of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2559 of SEQ ID NO:32. In a further embodiment of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:35. In yet another embodiment of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2502 of SEQ ID NO:34. In still another embodiment of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:37. In a particular embodiment of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2349 of SEQ ID NO:36. In an additional embodiment of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:39.

In one aspect of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2559 of SEQ ID NO:38. In another aspect of the invention

the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:41. In a further aspect of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2697 of SEQ ID NO:40. In yet another aspect of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:43. In still another aspect of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2583 of SEQ ID NO:42. In a particular aspect of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:45. In an additional aspect of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2397 of SEQ ID NO:44.

In one embodiment of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:47. In another embodiment of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2613 of SEQ ID NO:46. In a further embodiment of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:49. In yet another embodiment of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2619 of SEQ ID NO:48. In still another embodiment of the invention the gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:51. In a particular embodiment of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2790 of SEQ ID NO:50.

The invention thus provides a gene encodes a modified retinoblastoma protein comprising a contiguous amino acid sequence of SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:49 or SEQ ID NO:51. In one aspect of the invention the gene comprises a contiguous nucleic acid sequence from between position 7 and position 2691 of SEQ ID NO:28, from between position 7 and position 2628 of SEQ ID NO:30, from between position 7 and position 2559 of SEQ ID NO:32, from between position 7 and position 2502 of

SEQ ID NO:34, from between position 7 and position 2349 of SEQ ID NO:36, from between position 7 and position 2559 of SEQ ID NO:38, from between position 7 and position 2697 of SEQ ID NO:40, from between position 7 and position 2583 of SEQ ID NO:42, from between position 7 and position 2397 of SEQ ID NO:44, from between position 7 and position 2613 of SEQ ID NO:46, from between position 7 and position 2619 of SEQ ID NO:48 or from between position 7 and position 2790 of SEQ ID NO:50.

Another embodiment of the invention provides a DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴ or pRB⁵⁶, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification, where the DNA segment is operationally positioned under the control of a promoter. In one embodiment of the invention this DNA segment is operationally positioned under the control of a recombinant promoter. In another embodiment of the invention the DNA segment is further defined as a recombinant vector. In a particular aspect of the present invention, the recombinant vector is an adenoviral vector. In another aspect, the recombinant vector is a retroviral vector.

In a further embodiment of the invention the DNA segment is further defined as a component of a tetracycline responsive expression system. In yet another embodiment of the invention the DNA segment is operatively positioned downstream of a promoter comprising a tetracycline operator nucleic acid sequence; the tetracycline responsive expression system further comprising a second sequence region comprising an isolated gene encoding a fusion protein comprising a transcriptional transactivation domain operatively attached to a tetracycline repressor protein, the second sequence region operatively positioned downstream of a minimal promoter.

In yet another embodiment of the invention the tetracycline responsive expression system is comprised within an adenoviral vector. In still another embodiment of the invention the adenoviral vector is comprised within a recombinant adenovirus.

The invention also provides a DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification, which is comprised within a host cell. In one embodiment of the invention the host cell is a prokaryotic cell. In another
5 embodiment of the invention the host cell is a eukaryotic cell. In a further embodiment of the invention the host cell is a human cell. In yet another embodiment of the invention the host cell is a tumor cell. In still another embodiment of the invention the host cell is comprised within an animal. In a particular embodiment of the invention the animal is a human subject.

10 Another embodiment of the invention provides a DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification, which is dispersed in a pharmaceutically acceptable excipient.

Yet another embodiment of the invention provides an isolated DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification, wherein the modified retinoblastoma tumor suppressor protein is characterized as: comprising an N-terminal region that comprises at least a first sequence region from which at least one amino
15 acid has been deleted, and wherein the modified retinoblastoma tumor suppressor protein has a biological activity at least about equivalent to the biological activity of the corresponding wild-type retinoblastoma tumor suppressor protein; or comprising an N-terminal region that comprises a first sequence region comprising at least one mutation, and wherein the modified retinoblastoma tumor suppressor protein has an increased biological activity in comparison to the
20 biological activity of the corresponding wild-type retinoblastoma tumor suppressor protein.

In certain aspects of the invention, the DNA segments as described above are contemplated for use in expressing a modified retinoblastoma tumor suppressor protein, for example in a host cell. In other aspects, the DNA segments are contemplated for use in
25 inhibiting cellular proliferation, or in the preparation of a medicament for inhibiting cellular

proliferation or treating cancer, for example in a human patient. Thus, the use of the instant DNA segments in the preparation of a modified retinoblastoma tumor suppressor protein, in inhibiting cellular proliferation, and in the preparation of a medicament for inhibiting cellular proliferation or treating cancer is provided. In certain uses, the medicament is intended for administration to a human patient, or formulated for parenteral administration.

The invention further provides a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification.

The invention also provides a recombinant host cell comprising a DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification. In one aspect of the invention the host cell is a prokaryotic host cell. In another aspect of the invention the host cell is *E. coli*. In a further aspect of the invention the host cell is a eukaryotic host cell. In yet another aspect of the invention the host cell is a tumor cell. In still another aspect of the invention the DNA segment is introduced into the cell by means of a recombinant vector.

The invention further provides a method of inhibiting cellular proliferation, comprising contacting a cell with an effective inhibitory amount of a first modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, the modified retinoblastoma tumor suppressor protein comprising an N-terminal modification. In one embodiment of the invention the first modified retinoblastoma tumor suppressor protein comprises a modified retinoblastoma protein from which amino acids 111 through 241 have been deleted. In another embodiment of the invention the first modified retinoblastoma tumor suppressor protein comprises a modified retinoblastoma protein that comprises a mutation at position 111 and position 112. In a further embodiment of the invention the first modified retinoblastoma tumor suppressor protein is prepared by expressing a DNA segment encoding the modified retinoblastoma tumor suppressor protein in a recombinant host cell and collecting the modified retinoblastoma tumor suppressor protein

expressed by the cell. In yet another embodiment of the invention the cell is contacted with the first modified retinoblastoma tumor suppressor protein by providing to the cell a DNA segment that expresses the first modified retinoblastoma tumor suppressor protein in the cell. In still another embodiment of the invention the cell is provided with a tetracycline responsive expression vector system that expresses the first modified retinoblastoma tumor suppressor protein in the cell. In a particular embodiment of the invention the vector system is an adenoviral vector system.

Another aspect of the invention provides a method of inhibiting cellular proliferation, comprising contacting a tumor cell with an effective inhibitory amount of a first modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, the protein comprising an N-terminal modification. In one aspect of the invention the cell is located within an animal and the first modified retinoblastoma tumor suppressor protein, or a gene encoding the modified retinoblastoma tumor suppressor protein, is administered to the animal in a pharmaceutically acceptable vehicle. As used herein, the term "gene" is defined as an isolated DNA segment that includes the coding region of the protein, or a portion thereof. Thus the term "gene" includes genomic DNA, cDNA or RNA encoding the protein.

In another aspect of the invention the animal is a human subject. In a further aspect of the invention the cell is further contacted with a second tumor suppressor protein. In yet another aspect of the invention the cell is contacted with a modified retinoblastoma protein and a wild-type retinoblastoma, p53 or other tumor suppressor protein.

The invention further provides a method of inhibiting cellular proliferation, comprising contacting a cell with a retinoblastoma protein and a p53 protein in a combined amount effective to inhibit cellular proliferation in the cell.

The invention also provides a method of treating cancer, comprising administering to an animal with cancer a pharmaceutically acceptable composition comprising a biologically

effective inhibitory amount of a first modified retinoblastoma tumor suppressor protein, other than pRB⁹⁴, that comprises an N-terminal modification.

The terms "cancer" or "tumor" are clinically descriptive terms which encompass a myriad of diseases characterized by cells that exhibit unchecked and abnormal cellular proliferation. The term "tumor", when applied to tissue, generally refers to any abnormal tissue growth, *i.e.*, excessive and abnormal cellular proliferation. A tumor may be "benign" and unable to spread from its original focus, or "malignant" and capable of spreading beyond its anatomical site to other areas throughout the hostbody. The term "cancer" is an older term which is generally used to describe a malignant tumor or the disease state arising therefrom. Alternatively, the art refers to an abnormal growth as a neoplasm, and to a malignant abnormal growth as a malignant neoplasm.

Irrespective of whether the growth is classified as malignant or benign, the causes of excessive or abnormal cellular proliferation of tumor or cancer cells are not completely clear. Nevertheless, there is persuasive evidence that abnormal cellular proliferation is the result of a failure of one or more of the mechanisms controlling cell growth and division. It is also now believed that the mechanisms controlling cell growth and division include the genetic and tissue-mediated regulation of cell growth, mitosis and differentiation. These mechanisms are thought to act at the cell nucleus, the cell cytoplasm, the cell membrane and the tissue-specific environment of each cell. The process of transformation of a cell from a normal state to a condition of excessive or abnormal cellular proliferation is called tumorigenesis.

It has been observed that tumorigenesis is usually a multistep progression from a normal cellular state to, in some instances, a full malignancy. It is therefore believed that multiple "hits" upon the cell regulatory mechanisms are required for full malignancy to develop. Thus, in most instances, it is believed that there is no single cause of excessive proliferation, but that these disorders are the end result of a series of cumulative events.

While a malignant tumor or cancer capable of unchecked and rapid spread throughout the body is the most feared and usually the deadliest type of tumor, even so-called benign tumors or growths can cause significant morbidity and mortality by their inappropriate growth. A benign tumor can cause significant damage and disfigurement by inappropriate growth in cosmetically sensitive areas, or by exerting pressure on central or peripheral nervous tissue, blood vessels and other critical anatomical structures.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1. Relative activities of the modified hCMV promoters. The 5637 bladder carcinoma cells (lanes 1-5) and Saos2 osteocarcinoma cells (lanes 6-10) were transfected with reporter plasmids in which CAT gene expression was driven by the various modified (mhCMVp3, lanes 2 and 7; mhCMVp2, lanes 3 and 8; mhCMVp1, lanes 4 and 9) or full-length hCMV promoters (lanes 5 and 10). The % CAT activity is shown on the vertical axis. The CAT activity of the cells transfected with the plasmid carrying the full-length hCMV promoter (lanes 5 and 10) is defined as 100 percent.

FIG. 2. Expression of *tTA* from the modified mCMVp-*tTA* cassette has no squelching effects on the 5637 cell growth. A method of staining cells with crystal violet followed by measuring OD₅₅₀ was used for quantification of relative cell numbers (OD₅₅₀ shown on vertical axis; Gillies *et al.*, 1986). Shown is the growth parent cells with (▲) and without (□) tetracycline, and the mCMVp-*tTA* transfected cells with (◆) and without (○) tetracycline. Days after transfection are shown on the horizontal axis.

FIG. 3A, FIG. 3B and FIG. 3C. The effects of tetracycline-regulatable pRB expression on tumor cell growth (OD₅₅₀; vertical axis). **FIG. 3A.** Representative long-term clone from the *RB*-reconstituted osteosarcoma cell line (Saos-2, clone 11). **FIG. 3B.** Representative long-term clone from the *RB*-reconstituted breast carcinoma cell line (MDA-MB-468, clone 19-4). **FIG. 3C.** Representative long-term clone from the *RB*-reconstituted bladder carcinoma cell line (5637, clone 34-6). The cells were grown in the presence of 0.5 µg/ml of Tc (□) versus absence of Tc (○). Cell growth of the tumor cells stopped 1 to 2 days after pRB expression was turned on in Tc-free medium (days shown on horizontal axis). The growth cessation was irreversible at day 4 (arrows) after stimulation with fresh medium containing 15% serum (Saos-2), 10% serum plus 2 µg/ml phytohemagglutinin (PHA; MDA-468) or 10% serum plus 4 µg/ml of concanavalin A (Con A; 5637).

FIG. 4A, FIG. 4B and FIG. 4C. The effects of tetracycline-regulatable pRB expression on soft agar colony formation. **FIG. 4A.** Percent colony formation (vertical axis) for three independent Saos2 osteosarcoma cell line clones (RB110 Cl4, lane 2; RB110 Cl11, lane 3; RB110 Cl13, lane 4) and the Saos2 parent strain (lane 1). **FIG. 4B.** Percent colony formation (vertical axis) for two independent MDA-MB-468 breast carcinoma cell line clones (Rb110 Cl19-4, lane 2; Rb110Cl20-1, lane 3) and the MDA-MB-468 parent strain (lane 1). **FIG. 4C.** Percent colony formation (vertical axis) for two independent 5637 bladder carcinoma cell line clones (Rb110 Cl34-6, lane 2; Rb110 Cl36-9, lane 3) and the 5637 parent strain (lane 1). Soft agar colony formation of tumor cells with tetracycline-regulatable pRB expression was completely abrogated by induction of pRB in tetracycline-free medium. Colony formation is shown in the presence (open bar) and the absence (hatched bar) of tetracycline.

FIG. 5. Time course analysis of the pRB⁹⁴ and pRB¹¹⁰ expression in representative, Tc-regulatable Saos-2 cell clones in Tc-free media and its effects on DNA synthesis, using a ³H-thymidine incorporation assay. Lack of DNA synthesis as determined by failure of the tumor cells to incorporate thymidine implies growth cessation. The non-synchronized parental Saos-2 cell population (●) maintained steady DNA synthesis; Representative pRB¹¹⁰-reconstituted (■)

and pRB⁹⁴-reconstituted (♦) Saos-2 clones are illustrated. Percent ³H-labeled cells is shown on the vertical axis, and the hours after removal of tetracycline is shown on the horizontal axis.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

A. Tumor Suppressor Proteins

1. Retinoblastoma

Based upon study of the isolated RB cDNA clone, the predicted RB gene product has 928 amino acids and an expected molecular weight of 106 kDa (Lee *et al.*, 1987a; 1987b). The natural factor corresponding to the predicted RB gene expression product has been identified as a nuclear phosphoprotein having an apparent relative molecular mass (M_r) of between 105 and 114 kDa (Lee *et al.*, 1987b; Xu *et al.*, 1989b; Yokota *et al.*, 1988; Whyte *et al.*, 1988). The literature generally refers to the protein encoded by the RB gene as p110^{RB}. On SDS-PAGE normal human cells show an RB protein pattern consisting of a lower sharp band with an M_r of 110 kD and a broader, more variable region above this band with an M_r ranging from 110 kD to 116 kD. The 110 kD band is the underphosphorylated RB protein, whereas the broader region represents the phosphorylated RB protein. The heterogeneity of the molecular mass results from a varying degree of phosphorylation (Xu *et al.*, 1989b).

After years of intense scrutiny, the biological functions of the RB gene are beginning to be understood (reviewed in Cooper and Whyte, 1989; Hamel *et al.*, 1993; Horowitz, 1993; Riley *et al.*, 1994; Wang *et al.*, 1994; Weinberg, 1995). The RB protein shows cyclical changes in phosphorylation during the cell cycle. Most RB protein is unphosphorylated during G1 phase, but most (perhaps all) RB molecules are phosphorylated in S and G2 phases (Xu *et al.*, 1989b; DeCaprio *et al.*, 1989; Buchkovich *et al.*, 1989; Chen *et al.*, 1989; Mihara *et al.*, 1989). The established components of the pRB pathway include the E2F transcription factors, which are involved in transcriptional control of numerous cellular genes responsible for the advances of cells through the cell cycle (Nevins, 1992; La Thangue, 1994). The pRB also interacts with certain G1 phase cyclins (Koff *et al.*, 1992; Resnitzky and Reed, 1995; Geng *et al.*, 1996). Therefore, the RB gene apparently plays a key role in cell growth regulation being involved in

the major decisions during the G1 phase of the cell cycle which govern cell proliferation, quiescence and differentiation (Weinberg, 1995). Furthermore, only the underphosphorylated RB protein binds to SV40 large T antigen. Given that RB protein binding by large T antigen is probably important for the growth promoting effects of large T antigen, this suggests that the underphosphorylated RB protein is the active form of the RB protein, and the phosphorylated RB protein in S and G2 phases is inactive (Ludlow *et al.*, 1989).

It was reported that there was a striking difference in the ratio of underphosphorylated to phosphorylated pRB forms between normal fibroblasts growing exponentially and those arrested in G1 phase. More underphosphorylated pRB was observed in G1 arrested cells, suggesting the change in ratio of phosphorylated to underphosphorylated RB proteins was related to the fluctuation of cell cycle (Xu *et al.*, 1989b). Four subsequent papers have described the cell cycle-dependent phosphorylation of RB protein in detail (DeCaprio *et al.*, 1989; Buchkovich *et al.*; 1989; Chen *et al.*, 1989; Mihara *et al.*, 1989). It is now widely accepted that the product of the RB gene has a key role in the cell cycle control.

Cell proliferation depends on transcriptional activation of genes that are responsible for the onset of DNA synthesis as well as other critical events in the G1 phase of the cell cycle. As demonstrated by Pardee, transition of cells from a serum mitogens-dependent to serum mitogens-independent state is separated by a distinct time point at several hours before the onset of S phase, namely the R (restriction) point (Pardee, 1989). By passing through the R point, the cell commits itself to complete the remainder of the cell cycle through M phase. Therefore, the R point between the middle G1 and late G1 phases of the cell cycle represents a transition in the life of the cell that is as important as the G1/S boundary.

The phosphorylation status of pRB undergoes a readily distinguishable alteration at a time close to and perhaps contemporaneous with the R point transition of the cell cycle (Weinberg, 1995). During middle G1 phase, the only pRB species detected is an underphosphorylated form. When cells progress through the cell cycle, the pRB content increases gradually. However, the majority of pRB synthesized after middle G1 phase is

hyperphosphorylated. In other words, pRB hyperphosphorylation occurs in late G1, preceding the G1/S boundary (Xu *et al.*, 1991a; Mittnacht *et al.*, 1994). pRB maintains this hyperphosphorylated status throughout the remainder of the cell cycle, becoming dephosphorylated only upon evolution from M/early G1 (Ludlow *et al.*, 1990; Xu *et al.*, 1991a; Mittnacht *et al.*, 1994).

The underphosphorylated form of pRB is able to form complexes with the transcription factor E2Fs or directly interact with the E2F site, and switches the E2F site from a positive to negative element in transcriptional control. The E2F site is present in the promoters of diverse cellular genes that are responsible for the advances of cells through the cell cycle, including c-myc, B-myb, cdc2, dihydrofolate reductase, thymidine kinase, and the RB as well as the E2F-1 gene itself (Chellappan *et al.*, 1991; Nevins, 1992; Weintraub *et al.*, 1992; La Thangue, 1994; Shan *et al.*, 1994; Sardet *et al.*, 1995; Shan *et al.*, 1996). Since hyperphosphorylated pRB appears to have lost the ability to interact with E2Fs, the inhibitory function of pRB on cell growth can be abrogated by hyperphosphorylation.

The timing of pRB phosphorylation led to an attractive functional model (Weinberg, 1995). This model suggests that pRB is an R point guardian. pRB exerts most of its growth inhibitory effects in the first two thirds of the G1 phase. A cell that has progressed through early and middle G1 encounters the R point gate. Should conditions be ready for advance into the remainder of the cell cycle, pRB will undergo phosphorylation and functional inactivation, causing it to open the gate and to permit the cell to proceed into late G1. Cells that lack normal pRB function for various reasons will proceed freely into late G1. Without pRB, the upstream components of the cell cycle clock that regulate pRB phosphorylation, such as cyclin D, cyclin E and their corresponding cyclin-dependent kinases (CDKs) (Kato *et al.*, 1993; Ewen *et al.*, 1993) lose much of their influences in the decision of the cell to pass through the R point gate. Taken together, pRB allows the cell cycle clock to control the expression of numerous genes that mediate advance of the cell through a critical phase of its growth cycle being involved in the major decisions concurrent with the R point transition. Functional loss of pRB deprives the cell of this clock and thus of an important mechanism for braking cell proliferation.

Various mutations of the RB gene are known, and these are generally inactive. Mutations in RB are seen in virtually all cases of retinoblastoma; additionally, the RB gene products could potentially be inactivated by hyperphosphorylation, and by viral oncoprotein-like cellular protein binding. Although the RB gene was initially named because deletions or mutations within the gene caused the rare childhood ocular tumor, retinoblastoma, loss of pRB function is not only causally related to the retinoblastoma, but is also linked to the progression of many common human cancers. Additionally, there is growing evidence suggesting that the RB protein status is potentially a prognostic marker in urothelial carcinoma, non-small cell lung carcinoma, and perhaps also in some other types of human neoplasms (Xu, 1995).

In addition, with the revolutionary antigen retrieval technique and the available specific anti-pRB antibodies, immunohistochemistry has recently become one of the highly sensitive and reliable methods for detection of pRB inactivation in routinely processed pathological specimens (Xu, 1995). Altered pRB expression as determined by immunohistochemical analysis appears to signal a poor prognosis in a subset of human malignancies. It was initially reported that loss of functional pRB was a statistically significant negative prognostic factor in high-grade adult soft tissue sarcomas (Cance *et al.*, 1990). Subsequently, two independent studies done concurrently concluded that altered pRB expression was a prognostic factor among patients with transitional cell carcinoma of the bladder (Cordon-Cardo *et al.*, 1992; Logothetis *et al.*, 1992).

For lung cancer patients, the initial pilot studies have also been promising, implying that altered RB and p53 protein status could be a synergistic prognostic factor in early stage non-small cell lung carcinomas (Xu *et al.*, 1994a). A much worse survival pattern has been reported as well for patients with acute myelogenous leukemia who have low or absent levels of pRB protein in their peripheral blood leukemic cells (Kornblau *et al.*, 1994). Since all studies done so far to investigate association between the pRB status in human cancer and the clinical outcome of the patients have been retrospective, and the number of cases in each cohort was fairly small, definitive retrospective and prospective studies with an adequate sample size for statistical

calculations are now underway to determine whether or not loss of pRB function can be considered as a prognostic factor in clinical practice.

The most direct proof that the cloned RB gene is indeed a tumor suppressor gene comes from introduction of a cloned intact copy of the gene into cancer cells with observed tumor suppression function. A number of reports have indicated that replacement of the normal RB gene in RB-defective tumor cells from disparate types of human cancers could suppress their tumorigenic activity in nude mice (Huang *et al.*, 1988; Goodrich and Lee, 1993; Zhou *et al.*, 1994b). The tumor cell lines studied were derived from widely disparate types of human cancers such as the retinoblastoma, osteosarcoma, carcinomas of the bladder, prostate, breast and lung (Table 2).

Of note, there has been a tendency in the literature to separate the inhibition of cell growth by RB replacement in RB-defective tumor cells from its tumor suppression function (Takahashi *et al.*, 1991; Chen *et al.*, 1992; Goodrich *et al.*, 1992b; Zhou *et al.*, 1994b). After transient transduction with a wild-type pRB-expressing retrovirus or plasmid, as documented in several early studies, the RB-deficient retinoblastoma and osteosarcoma tumor cells in culture displayed striking changes, including cell enlargement, senescent phenotype and lower growth rate (Huang *et al.*, 1988; Templeton *et al.*, 1991). Subsequently, it was found that long-term stable clones of the RB-reconstituted tumor cells can be isolated that grew just as rapidly as the parental or matched RB⁻ revertant clones. The majority of RB⁺ clones obtained, however, were non-tumorigenic or with significantly reduced tumorigenicity in nude mice. The mechanisms for the dissociation of suppression of tumorigenicity in nude mice from inhibition of tumor cell growth in culture by RB-replacement are unclear. It is certainly possible that RB replacement restores sensitivity to a variety of physiologic growth inhibitory signals which may be present and supplied to cells when tumorigenicity assay is done in nude mice. Such external growth inhibitory agents would be absent under regular cell culture conditions, leading to rapid cell growth (Chen *et al.*, 1992).

Although the molecular mechanism of the RB-mediated tumor suppression have remained unclear, suppression of tumorigenicity of RB⁻ tumor cells *in vivo* by re-expressing the wild-type pRB implies that the RB gene could be a potential therapeutic target for human cancer. In addition, recent reports suggest that RB may also play a role in elicitation of immunogenicity of tumor cells (Lu *et al.*, 1994; Lu *et al.*, 1996), anti-angiogenesis (Dawson *et al.*, 1995) and suppression of tumor invasiveness (Li *et al.*, 1996), which make the emerging RB gene therapy even more attractive. In this regard, preclinical studies have recently demonstrated that treatment of established human xenograft tumors in nude mice by recombinant adenovirus vectors expressing either wild-type or an N-terminal truncated retinoblastoma protein resulted in regression of the treated tumors (Xu *et al.*, 1996). In addition, a constitutively active form of the pRB protein has been tested in a rat artery model of restenosis to inhibit vascular proliferative disorders following balloon angioplasty (Chang *et al.*, 1995).

The RB gene expressing the first in-frame AUG codon-initiated RB protein is also referred to herein as the intact RB gene, the RB¹¹⁰ gene or the p110^{RB} coding gene. It has also been observed that lower molecular weight (<100 kD, 98 kD, or 98-104 kD) bands of unknown origin which are immunoreactive to various anti-RB antibodies can be detected in immunoprecipitation and Western blots (Xu *et al.*, 1989b; Furukawa *et al.*, 1990; Stein *et al.*, 1990).

The RB¹¹⁰ cDNA open reading frame sequence (McGee *et al.*, 1989) contains a second in-frame AUG codon located in exon 3, at nucleotides 355-357. The deduced second AUG codon-initiated RB protein would be 98 kD, or 12 kD smaller than the p110^{RB} protein. It has been proposed that the lower molecular weight bands are the underphosphorylated (98 kD) and phosphorylated (98-104 kD) RB protein translated from the second AUG codon of the RB mRNA (Xu *et al.*, 1989b), and this was later shown conclusively (Xu *et al.*, U.S. Patent 5,496,731). This protein is referred to as the p94^{RB} protein.

It has been proposed that introduction of a functional RB¹¹⁰ gene into an RB-minus tumor cell will likely "normalize" the cell. Of course, it was not expected that tumor cells which

already have normal RB¹¹⁰ gene expression ("RB⁺") would respond to RB¹¹⁰ gene therapy, because it was presumed that adding additional RB expression could not correct a non-RB genetic defect. In fact, it has been shown that in the case of RB⁺ tumor cell lines, such as the osteosarcoma cell line U-2 OS, which expresses the normal p110^{RB}, introduction of an extra p110^{RB} coding gene did not change the neoplastic phenotype of such tumor lines (Huang *et al.*, 1988).

In the only reported exception, introduction of a p110^{RB} coding vector into normal human fibroblasts, WS1, which have no known RB or any other genetic defects, led to the cessation of cell growth (Fung *et al.*, WO 91/15580, 1991). However, it is believed that these findings were misinterpreted since a plasmid, ppVUO-Neo, producing SV40 T antigen with a well-known growth-promoting effect on host cells was used improperly to provide a comparison with the effect of RB¹¹⁰ expression on cell growth of transfected WS1 fibroblasts (Fung *et al.* WO 91/15580, 1991). This view is confirmed by the extensive literature, clearly characterizing RB⁺ tumor cells as "incurable" by treatment with wild-type RB¹¹⁰ gene. In addition, it is noteworthy that the WS1 cell line per se is a generally recognized non-tumorigenic human diploid fibroblast cell line with limited cell division potential in culture. Therefore, WO91/15580 simply does not provide any method for effectively treating RB⁺ tumors with an RB¹¹⁰ gene. Thus, there remains a need for a broad-spectrum tumor suppressor gene for treating abnormally proliferating cells having any type of genetic defect.

2. p53

Somatic cell mutations of the p53 gene are said to be the most frequently mutated gene in human cancer (Weinberg, 1991). The normal or wild-type p53 gene is a negative regulator of cell growth, which, when damaged, favors cell transformation (Weinberg, 1991). As noted for the RB protein, the p53 expression product is found in the nucleus, where it may act in parallel with or cooperatively with p110^{RB}. This is suggested by a number of observations, for example, both p53 and p110^{RB} proteins are targeted for binding or destruction by the oncoproteins of SV40, adenovirus and human papillomavirus. Tumor cell lines deleted for p53 have been successfully treated with wild-type p53 vector to reduce tumorigenicity (Baker *et al.*, 1990).

However, the introduction of either p53 or RB¹¹⁰ into cells that have not undergone lesions at these loci does not affect cell proliferation (Marshall, 1991; Baker *et al.*, 1990; Huang *et al.*, 1988). Such experiments suggest that sensitivity of cells to the suppression of their growth by a tumor suppressor gene is dependent on the genetic alterations that have taken place in the cells. Such a dependency would be further complicated by the observation in certain cancers that alterations in the p53 tumor suppressor or gene locus appear after mutational activation of the ras oncogene (Marshall, 1991; Fearon *et al.*, 1990a). Therefore, there remains a need for a broad-spectrum tumor suppressor gene that does not depend on the specific identification of each mutated gene causing abnormal cellular proliferation.

3. Neurofibromatosis Type 1

Neurofibromatosis type 1 or von Recklinghausen neurofibromatosis results from the inheritance of a predisposing mutant allele or from alleles created through new germline mutations (Marshall, 1991). The neurofibromatosis type 1 gene, referred to as the NF1 gene, is a relatively large locus exhibiting a mutation rate of around 10^{-4} . Defects in the NF1 gene result in a spectrum of clinical syndromes ranging from cafe-au-lait spots to neurofibromas of the skin and peripheral nerves to Schwannomas and neurofibrosarcomas. The NF1 gene encodes a protein of about 2485 amino acids that shares structural similarity with three proteins that interact with the products of the ras protooncogene (Weinberg, 1991). For example, the NF1 amino acid sequence shows sequence homology to the catalytic domain of ras GAP, a GTPase-activating protein for p21 ras (Marshall, 1991).

The role of NF1 in cell cycle regulation is apparently a complex one that is not yet fully elucidated. For example, it has been hypothesized that it is a suppressor of oncogenically activated p21 ras in yeast (Marshall, 1991 citing Ballester *et al.*, 1990). On the other hand, other possible pathways for NF1 interaction are suggested by the available data (Marshall, 1991; Weinberg, 1991). At present, no attempts to treat NF1 cells with a wild-type NF1 gene have been undertaken due to the size and complexity of the NF1 locus. Therefore, it would be highly desirable to have a broad-spectrum tumor suppressor gene able to treat NF1 and any other type of cancer or tumor.

4. DCC

The multiple steps in the tumorigenesis of colon cancer are readily monitored during development by colonoscopy. The combination of colonoscopy with the biopsy of the involved tissue has uncovered a number of degenerative genetic pathways leading to the result of a malignant tumor. One well studied pathway begins with large polyps in which 60% of the cells carry a mutated, activated allele of K-ras. A majority of these tumors then proceed to the inactivation-mutation of the gene referred to as the deleted in colon carcinoma (DCC) gene, followed by the inactivation of the p53 tumor suppressor gene.

The DCC gene is a more than approximately one million base pair gene coding for a 190-kD transmembrane phosphoprotein which is hypothesized to be a receptor (Weinberg, 1991), the loss of which allows the affected cell a growth advantage. It has also been noted that the DCC has partial sequence homology to the neural cell adhesion molecule (Marshall, 1991) which might suggest a role for the DCC protogene in regulating cell to cell interactions. As can be appreciated, the large size and complexity of the DCC gene, together with the complexity of the K-ras, p53 and possibly other genes involved in colon cancer tumorigenesis demonstrates a need for a broad-spectrum tumor suppressor gene and methods of treating colon carcinoma cells which do not depend upon manipulation of the DCC gene or on the identification of other specific damaged genes in colon carcinoma cells.

5. Other Tumor Suppressor Proteins

Examples of additional tumor suppressor genes and candidate tumor suppressor genes contemplated for use in combination with the tumor suppressor genes of the present invention include, but are not limited to; the Wilms tumor (WT-1) gene (Call *et al.*, 1990; Gessler *et al.*, 1990; Pritchard-Jones *et al.*, 1990), the von Hippel-Lindau (VHL) disease tumor suppressor gene (Duan *et al.*, 1995), the Maspin (Zou *et al.*, 1994), Brush-1 (Schott *et al.*, 1994) and BRCA 1 genes (Miki *et al.*, 1994; Futreal *et al.*, 1994) for breast cancer, and the multiple tumor suppressor (MTS) or p16 gene (Serrano *et al.*, 1993; Kamb *et al.*, 1994).

B. DNA Delivery *via* Infection with Viral Vectors

In certain embodiments of the invention, the tumor suppressor genes may be stably integrated into the genome of the cell. In yet further embodiments, the genes may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance or replication independent of or in synchronization with the host cell cycle. How the tumor suppressor gene is delivered to a cell and where in the cell the nucleic acid remains is dependent on the type of expression vector employed.

1. Adenoviral Vectors

Preferred for use in the present invention are adenovirus vectors, and particularly tetracycline-controlled adenovirus vectors. These vectors may be employed to deliver and express a wide variety of genes, including, but not limited to, tumor suppressor genes such as the retinoblastoma and p53 genes, in addition to cytokine genes such as tumor necrosis factor α , the interferon gene family and the interleukin gene family.

A preferred method for delivery of the expression constructs involves the use of an adenovirus expression vector. Although adenovirus vectors are known to have a low capacity for integration into genomic DNA, this feature is counterbalanced by the high efficiency of gene transfer afforded by these vectors. "Adenovirus expression vector" is meant to include those constructs containing adenovirus sequences sufficient to (a) support packaging of the construct in host cells with complementary packaging functions and (b) to ultimately express a heterologous gene of interest that has been cloned therein.

The expression vector comprises a genetically engineered form of adenovirus. Knowledge of the genetic organization of adenovirus, a 36 kb, linear, double-stranded DNA virus, allows substitution of large pieces of adenoviral DNA with foreign sequences (Grunhaus and Horwitz, 1992). In contrast to retrovirus, the adenoviral infection of host cells does not result in chromosomal integration because wild-type adenoviral DNA can replicate in an

episomal manner without potential genotoxicity. Also, adenoviruses are structurally stable, and no genome rearrangement has been detected after extensive amplification.

Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized genome, ease of manipulation, high titer, wide target-cell range and high infectivity. Both ends of the viral genome contain 100-200 base pair inverted repeats (ITRs), which are *cis* elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome contain different transcription units that are divided by the onset of viral DNA replication. The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression and host cell shut-off (Renan, 1990). The products of the late genes, including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary transcript issued by the major late promoter (MLP). The MLP, (located at 16.8 m.u.) is particularly efficient during the late phase of infection, and all the mRNA's issued from this promoter possess a 5'-tripartite leader (TPL) sequence which makes them preferred mRNA's for translation.

In a current system, recombinant adenovirus is generated from homologous recombination between a shuttle vector and a master plasmid which contains the backbone of the adenovirus genome. Due to the possible recombination between the backbone of the adenovirus genome, and the cellular DNA of the helper cells which contain the missing portion of the viral genome, wild-type adenovirus may be generated from this process. Therefore, it is critical to isolate a single clone of virus from an individual plaque and examine its genomic structure.

Generation and propagation of most adenovirus vectors, which are replication deficient, depend on a unique helper cell line, designated 293, which was transformed from human embryonic kidney cells by Ad5 DNA fragments and constitutively expresses E1 proteins (E1A and E1B; Graham *et al.*, 1977). Since the E3 region is dispensable from the adenovirus genome (Jones and Shenk, 1978), the current adenovirus vectors, with the help of 293 cells, carry foreign

DNA in either the E1, the E3 or both regions (Graham and Prevec, 1991). In nature, adenovirus can package approximately 105% of the wild-type genome (Ghosh-Choudhury *et al.*, 1987), providing capacity for about 2 extra kb of DNA. Combined with the approximately 5.5 kb of DNA that is replaceable in the E1 and E3 regions, the maximum capacity of most adenovirus vectors is at least 7.5 kb, or about 15% of the total length of the vector. More than 80% of the adenovirus viral genome remains in the vector backbone.

Gene transfer *in vivo* using recombinant E1-deficient adenoviruses results in early and late viral gene expression that may elicit a host immune response, thereby limiting the duration of transgene expression and the use of adenoviruses for gene therapy. In order to circumvent these potential problems, the prokaryotic Cre-loxP recombination system has been adapted to generate recombinant adenoviruses with extended deletions in the viral genome in order to minimize expression of immunogenic and/or cytotoxic viral proteins (Lieber *et al.*, 1996).

Helper cell lines may be derived from human cells such as human embryonic kidney cells, muscle cells, hematopoietic cells or other human embryonic mesenchymal or epithelial cells. Alternatively, the helper cells may be derived from the cells of other mammalian species that are permissive for human adenovirus. Such cells include, *e.g.*, Vero cells or other monkey embryonic mesenchymal or epithelial cells. As stated above, the preferred helper cell line is 293.

Recently, Racher *et al.* (1995) disclosed improved methods for culturing 293 cells and propagating adenovirus. In one format, natural cell aggregates are grown by inoculating individual cells into 1 liter siliconized spinner flasks (Techne, Cambridge, UK) containing 100-200 ml of medium. Following stirring at 40 rpm, the cell viability is estimated with trypan blue. In another format, Fibra-Cel microcarriers (Bibby Sterlin, Stone, UK) (5 g/l) is employed as follows. A cell inoculum, resuspended in 5 ml of medium, is added to the carrier (50 ml) in a 250 ml Erlenmeyer flask and left stationary, with occasional agitation, for 1 to 4 h. The medium is then replaced with 50 ml of fresh medium and shaking initiated. For virus production, cells are allowed to grow to about 80% confluence, after which time the medium is replaced (to 25% of the final volume) and adenovirus added at an MOI of 0.05. Cultures are left stationary

overnight, following which the volume is increased to 100% and shaking commenced for another 72 h.

In some cases, adenovirus mediated gene delivery to multiple cell types has been found to be much less efficient compared to epithelial derived cells. A new adenovirus, AdPK, has been constructed to overcome this inefficiency (Wickham *et al.*, 1996). AdPK contains a heparin-binding domain that targets the virus to heparan-containing cellular receptors, which are broadly expressed in many cell types. Therefore, AdPK delivers genes to multiple cell types at higher efficiencies than unmodified adenovirus, thus improving gene transfer efficiency and expanding the tissues amenable to efficient adenovirus mediated gene therapy.

Other than the requirement that the adenovirus vector be replication defective, or at least conditionally defective, the nature of the adenovirus vector is not believed to be crucial to the successful practice of the invention. The adenovirus may be of any of the 42 different known serotypes or subgroups A-F. Adenovirus type 5 of subgroup C is the preferred starting material in order to obtain the conditional replication-defective adenovirus vector for use in the present invention. This is because Adenovirus type 5 is a human adenovirus about which a great deal of biochemical and genetic information is known, and it has historically been used for most constructions employing adenovirus as a vector.

As stated above, the typical vector according to the present invention is replication defective and will not have an adenovirus E1 region. Thus, it will be most convenient to introduce the foreign gene expression cassette at the position from which the E1-coding sequences have been removed. However, the position of insertion of the construct within the adenovirus sequences is not critical to the invention. The polynucleotide encoding the gene of interest may also be inserted in lieu of the deleted E3 region in E3 replacement vectors as described by Karlsson *et al.* (1986) or in the E4 region where a helper cell line or helper virus complements the E4 defect (Brough *et al.*, 1996).

Adenovirus growth and manipulation is known to those of skill in the art, and exhibits broad host range *in vitro* and *in vivo*. This group of viruses can be obtained in high titers, *e.g.*, 10^9 to 10^{11} plaque-forming units per ml, and they are highly infective. The life cycle of adenovirus does not require integration into the host cell genome. The foreign genes delivered
5 by adenovirus vectors are episomal and, therefore, have low genotoxicity to host cells. No severe side effects have been reported in studies of vaccination with wild-type adenovirus (Couch *et al.*, 1963; Top *et al.*, 1971), demonstrating their safety and therapeutic potential as *in vivo* gene transfer vectors.

Adenovirus vectors have been used in eukaryotic gene expression (Levrero *et al.*, 1991; Gomez-Foix *et al.*, 1992) and vaccine development (Grunhaus and Horwitz, 1992; Graham and Prevec, 1992). Recently, animal studies suggested that recombinant adenovirus could be used for gene therapy (Stratford-Perricaudet and Perricaudet, 1991; Stratford-Perricaudet *et al.*, 1991; Rich *et al.*, 1993). Studies in administering recombinant adenovirus to different tissues include trachea instillation (Rosenfeld *et al.*, 1991; 1992), muscle injection (Ragot *et al.*, 1993), peripheral intravenous injections (Herz and Gerard, 1993) and stereotactic inoculation into the brain (Le Gal La Salle *et al.*, 1993). Recombinant adenovirus and adeno-associated virus (see below) can both infect and transduce non-dividing human primary cells.

2. AAV Vectors

Adeno-associated virus (AAV) is also an attractive system for use in construction of vectors for delivery of and expression of tumor suppressor genes as it has a high frequency of integration and it can infect nondividing cells, thus making it useful for delivery of genes into mammalian cells, for example, in tissue culture (Muzyczka, 1992) or *in vivo*. AAV has a broad
25 host range for infectivity (Tratschin *et al.*, 1984; Laughlin *et al.*, 1986; Lebkowski *et al.*, 1988; McLaughlin *et al.*, 1988). Details concerning the generation and use of rAAV vectors are described in U.S. Patent No. 5,139,941 and U.S. Patent No. 4,797,368, each incorporated herein by reference.

Studies demonstrating the use of AAV in gene delivery include LaFace *et al.* (1988); Zhou *et al.* (1993); Flotte *et al.* (1993); and Walsh *et al.* (1994). Recombinant AAV vectors have been used successfully for *in vitro* and *in vivo* transduction of marker genes (Kaplitt *et al.*, 1994; Lebkowski *et al.*, 1988; Samulski *et al.*, 1989; Yoder *et al.*, 1994; Zhou *et al.*, 1994a; Hermonat and Muzyczka, 1984; Tratschin *et al.*, 1985; McLaughlin *et al.*, 1988) and genes involved in human diseases (Flotte *et al.*, 1992; Luo *et al.*, 1994; Ohi *et al.*, 1990; Walsh *et al.*, 1994; Wei *et al.*, 1994). Recently, an AAV vector has been approved for phase I human trials for the treatment of cystic fibrosis.

AAV is a dependent parvovirus in that it requires coinfection with another virus (either adenovirus or a member of the herpes virus family) to undergo a productive infection in cultured cells (Muzyczka, 1992). In the absence of coinfection with helper virus, the wild type AAV genome integrates through its ends into human chromosome 19 where it resides in a latent state as a provirus (Kotin *et al.*, 1990; Samulski *et al.*, 1991). rAAV, however, is not restricted to chromosome 19 for integration unless the AAV Rep protein is also expressed (Shelling and Smith, 1994). When a cell carrying an AAV provirus is superinfected with a helper virus, the AAV genome is "rescued" from the chromosome or from a recombinant plasmid, and a normal productive infection is established (Samulski *et al.*, 1989; McLaughlin *et al.*, 1988; Kotin *et al.*, 1990; Muzyczka, 1992).

Typically, recombinant AAV (rAAV) virus is made by cotransfecting a plasmid containing the gene of interest flanked by the two AAV terminal repeats (McLaughlin *et al.*, 1988; Samulski *et al.*, 1989; each incorporated herein by reference) and an expression plasmid containing the wild type AAV coding sequences without the terminal repeats, for example pIM45 (McCarty *et al.*, 1991; incorporated herein by reference). The cells are also infected or transfected with adenovirus or plasmids carrying the adenovirus genes required for AAV helper function. rAAV virus stocks made in such fashion are contaminated with adenovirus which must be inactivated by heat shock or physically separated from the rAAV particles (for example, by cesium chloride density centrifugation). Alternatively, adenovirus vectors containing the AAV coding regions or cell lines containing the AAV coding regions and some or all of the adenovirus

helper genes could be used (Yang *et al.*, 1994; Clark *et al.*, 1995). Cell lines carrying the rAAV DNA as an integrated provirus can also be used (Flotte *et al.*, 1995).

3. Retroviral Vectors

In particular aspects of the present invention, delivery of selected genes to target cells through the use of retrovirus infection will be desired. The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of reverse-transcription (Coffin, 1990). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins. The integration results in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes, gag, pol, and env that code for capsid proteins, polymerase enzyme, and envelope components, respectively. A sequence found upstream from the gag gene contains a signal for packaging of the genome into virions. Two long terminal repeat (LTR) sequences are present at the 5' and 3' ends of the viral genome. These contain strong promoter and enhancer sequences and are also required for integration in the host cell genome (Coffin, 1990).

In order to construct a retroviral vector, a nucleic acid encoding a gene of interest is inserted into the viral genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol, and env genes but without the LTR and packaging components is constructed (Mann *et al.*, 1983). When a recombinant plasmid containing a cDNA, together with the retroviral LTR and packaging sequences is introduced into this cell line (by calcium phosphate precipitation for example), the packaging sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). The media containing the recombinant retroviruses is then collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression require the division of host cells (Paskind *et al.*, 1975).

Concern with the use of defective retrovirus vectors is the potential appearance of wild-type replication-competent virus in the packaging cells. This can result from recombination events in which the intact sequence from the recombinant virus inserts upstream from the gag, pol, env sequence integrated in the host cell genome. However, new packaging cell lines are now available that should greatly decrease the likelihood of recombination (Markowitz *et al.*, 1988; Hersdorffer *et al.*, 1990).

In some cases, the restricted host-cell range and low titer of retroviral vectors can limit their use for stable gene transfer in eukaryotic cells. To overcome these potential difficulties, a murine leukemia virus-derived vector has been developed in which the retroviral envelope glycoprotein has been completely replaced by the G glycoprotein of vesicular stomatitis virus (Burns *et al.*, 1993). These vectors can be concentrated to extremely high titers (10^9 colony forming units/ml), and can infect cells that are ordinarily resistant to infection with vectors containing the retroviral envelope protein. These vectors may facilitate gene therapy model studies and other gene transfer studies that require direct delivery of vectors *in vivo*.

4. Baculoviral Vectors

Baculovirus expression vectors are useful tools for the production of proteins for a variety of applications (Summers and Smith, 1987; O'Reilly *et al.*, 1992; also U.S. Patent Nos., 4,745,051 (Smith and Summers), 4,879,236 (Smith and Summers), 5,077,214 (Guarino and Jarvis), 5,155,037 (Summers), 5,162,222, (Guarino and Jarvis), 5,169,784 (Summers and Oker-Blom) and 5,278,050 (Summers), each incorporated herein by reference). The inventors contemplate the construction of baculoviral expression vectors wherein gene expression is regulated by tetracycline. These vectors might be particularly useful, for example, where the desired protein is toxic to the insect cells. In these instances, production of the protein can be turned off until the cells have reached a very high density, thereby still allowing for the production of large quantities of the desired protein.

Baculovirus expression vectors are recombinant insect vectors in which the coding region of a particular gene of interest is placed behind a promoter in place of a nonessential baculoviral

gene. The classic approach used to isolate a recombinant baculovirus expression vector is to construct a plasmid in which the foreign gene of interest is positioned downstream of the *polyhedrin* promoter. Then, *via* homologous recombination, that plasmid can be used to transfer the new gene into the viral genome in place of the wild-type *polyhedrin* gene (Summers and Smith, 1987; O'Reilly *et al.*, 1992).

The resulting recombinant virus can infect cultured lepidopteran insect cells or larvae and express the foreign gene under the control of the *polyhedrin* promoter, which is strong and provides very high levels of transcription during the very late phase of infection. The strength of the *polyhedrin* promoter is an advantage of the use of recombinant baculoviruses as expression vectors because it usually leads to the synthesis of large amounts of the foreign gene product during infection.

5. Other viral vectors

Other viral vectors may be employed for construction of expression vectors in the present invention. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar *et al.*, 1988), sindbis virus and herpesviruses may be employed. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar *et al.*, 1988; Horwich *et al.*, 1990).

With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. *In vitro* studies showed that the virus could retain the ability for helper-dependent packaging and reverse transcription despite the deletion of up to 80% of its genome (Horwich *et al.*, 1990). This suggested that large portions of the genome could be replaced with foreign genetic material. Chang *et al.* (1991) recently introduced the chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B virus genome in the place of the polymerase, surface, and pre-surface coding sequences. It was cotransfected with wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang *et al.*, 1991).

6. Modified Viruses

In still further embodiments of the present invention, particularly wherein delivery of a selected gene to a specific cell type is desired, the expression constructs to be delivered are housed within an infective virus that has also been engineered to express a specific binding ligand. The virus particle will thus bind specifically to the cognate receptors of the target cell and deliver the contents to the cell. A novel approach designed to allow specific targeting of retrovirus vectors was recently developed based on the chemical modification of a retrovirus by the chemical addition of lactose residues to the viral envelope. This modification can permit the specific infection of hepatocytes via sialoglycoprotein receptors.

Another approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled *via* the biotin components by using streptavidin (Roux *et al.*, 1989). Using antibodies against major histocompatibility complex class I and class II antigens, they demonstrated the infection of a variety of human cells that bore those surface antigens with an ecotropic virus *in vitro* (Roux *et al.*, 1989).

C. Other Methods of DNA Delivery

As well as the viral mediated methods of DNA delivery *via* infection of cells described above, other methods of introducing the tumor suppressor genes of the present invention into both prokaryotic and eukaryotic cells are contemplated.

1. Transfection and Transformation

In order to effect expression of a gene construct, the expression construct must be delivered into a cell. As described herein, a preferred mechanism for delivery is *via* viral infection, where the expression construct is encapsidated in an infectious viral particle. However, several non-viral methods for the transfer of expression constructs into eukaryotic and prokaryotic cells also are contemplated by the present invention. In one embodiment of the present invention, the expression construct may consist only of naked recombinant DNA or

plasmids. Transfer of the construct may be performed by any of the methods mentioned which physically or chemically permeabilize the cell membrane.

a. Liposome-Mediated Transfection and Transformation

In a further embodiment of the invention, the expression construct may be entrapped in a liposome. Liposomes are vesicular structures characterized by a phospholipid bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers (Ghosh and Bachhawat, 1991). Also contemplated is an expression construct complexed with Lipofectamine (Gibco BRL).

Liposome-mediated nucleic acid delivery and expression of foreign DNA *in vitro* has been very successful (Nicolau and Sene, 1982; Fraley *et al.*, 1979; Nicolau *et al.*, 1987). Wong *et al.* (1980) demonstrated the feasibility of liposome-mediated delivery and expression of foreign DNA in cultured chick embryo, HeLa and hepatoma cells.

In certain embodiments of the invention, the liposome may be complexed with a hemagglutinating virus (HVJ). This has been shown to facilitate fusion with the cell membrane and promote cell entry of liposome-encapsulated DNA (Kaneda *et al.*, 1989). In other embodiments, the liposome may be complexed or employed in conjunction with nuclear non-histone chromosomal proteins (HMG-1) (Kato *et al.*, 1991). In yet further embodiments, the liposome may be complexed or employed in conjunction with both HVJ and HMG-1.

b. Electroporation

In certain embodiments of the present invention, the expression construct is introduced into the cell *via* electroporation. Electroporation involves the exposure of a suspension of cells and DNA to a high-voltage electric discharge.

Transfection of eukaryotic cells using electroporation has been quite successful. Mouse pre-B lymphocytes have been transfected with human kappa-immunoglobulin genes (Potter *et al.*, 1984), and rat hepatocytes have been transfected with the chloramphenicol acetyltransferase gene (Tur-Kaspa *et al.*, 1986) in this manner.

5

c. Calcium Phosphate Precipitation or DEAE-Dextran Treatment

In other embodiments of the present invention, the expression construct is introduced to the cells using calcium phosphate precipitation. Human KB cells have been transfected with adenovirus 5 DNA (Graham and Van Der Eb, 1973) using this technique. Also in this manner,
10 mouse L(A9), mouse C127, CHO, CV-1, BHK, NIH3T3 and HeLa cells were transfected with a neomycin marker gene (Chen and Okayama, 1987), and rat hepatocytes were transfected with a variety of marker genes (Rippe *et al.*, 1990).

In another embodiment, the expression construct is delivered into the cell using DEAE-dextran followed by polyethylene glycol. In this manner, reporter plasmids were introduced into mouse myeloma and erythroleukemia cells (Gopal, 1985).

d. Particle Bombardment

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA-coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter cells without killing them (Klein *et al.*, 1987). Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical current, which in turn provides the motive force (Yang *et al.*, 1990). The
20 microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

25

e. Direct Microinjection or Sonication Loading

Further embodiments of the present invention include the introduction of the expression
30 construct by direct microinjection or sonication loading. Direct microinjection has been used to

introduce nucleic acid constructs into *Xenopus* oocytes (Harland and Weintraub, 1985), and LTK⁻ fibroblasts have been transfected with the thymidine kinase gene by sonication loading (Fechheimer *et al.*, 1987).

f. Adenoviral Assisted Transfection

In certain embodiments of the present invention, the expression construct is introduced into the cell using adenovirus assisted transfection. Increased transfection efficiencies have been reported in cell systems using adenovirus coupled systems (Kelleher and Vos, 1994; Cotten *et al.*, 1992; Curiel, 1994).

g. Receptor Mediated Transfection

Still further expression constructs that may be employed to deliver the construct to the target cells are receptor-mediated delivery vehicles. These take advantage of the selective uptake of macromolecules by receptor-mediated endocytosis that will be occurring in the target cells. In view of the cell type-specific distribution of various receptors, this delivery method adds a degree of specificity to the present invention. Specific delivery in the context of another mammalian cell type is described by Wu and Wu (1993; incorporated herein by reference).

Certain receptor-mediated gene targeting vehicles comprise a cell receptor-specific ligand and a DNA-binding agent. Others comprise a cell receptor-specific ligand to which the DNA construct to be delivered has been operatively attached. Several ligands have been used for receptor-mediated gene transfer (Wu and Wu, 1987; Wagner *et al.*, 1990; Perales *et al.*, 1994; Myers, EPO 0273085), which establishes the operability of the technique. In the context of the present invention, the ligand will be chosen to correspond to a receptor specifically expressed on the neuroendocrine target cell population.

In other embodiments, the DNA delivery vehicle component of a cell-specific gene targeting vehicle may comprise a specific binding ligand in combination with a liposome. The nucleic acids to be delivered are housed within the liposome and the specific binding ligand is functionally incorporated into the liposome membrane. The liposome will thus specifically bind

to the receptors of the target cell and deliver the contents to the cell. Such systems have been shown to be functional using systems in which, for example, epidermal growth factor (EGF) is used in the receptor-mediated delivery of a nucleic acid to cells that exhibit upregulation of the EGF receptor.

5

In still further embodiments, the DNA delivery vehicle component of the targeted delivery vehicles may be a liposome itself, which will preferably comprise one or more lipids or glycoproteins that direct cell-specific binding. For example, Nicolau *et al.* (1987) employed lactosyl-ceramide, a galactose-terminal asialoganglioside, incorporated into liposomes and observed an increase in the uptake of the insulin gene by hepatocytes. It is contemplated that the tissue-specific transforming constructs of the present invention can be specifically delivered into the target cells in a similar manner.

10

D. Marker Genes

In certain aspects of the present invention, specific cells are tagged with specific genetic markers to provide information about the fate of the tagged cells. Therefore, the present invention also provides recombinant candidate screening and selection methods which are based upon whole cell assays and which, preferably, employ a reporter gene that confers on its recombinant hosts a readily detectable phenotype that emerges only under conditions where a general DNA promoter positioned upstream of the reporter gene is functional. Generally, reporter genes encode a polypeptide (marker protein) not otherwise produced by the host cell which is detectable by analysis of the cell culture, *e.g.*, by fluorometric, radioisotopic or spectrophotometric analysis of the cell culture.

20

In other aspects of the present invention, a genetic marker is provided which is detectable by standard genetic analysis techniques, such as DNA or RNA amplification by PCR™ or hybridization using fluorometric, radioisotopic or spectrophotometric probes.

25

1. Screening

Exemplary enzymes include esterases, phosphatases, proteases (tissue plasminogen activator or urokinase) and other enzymes capable of being detected by their activity, as will be known to those skilled in the art. Contemplated for use in the present invention is green fluorescent protein (GFP) as a marker for transgene expression (Chalfie *et al.*, 1994). The use of GFP does not need exogenously added substrates, only irradiation by near UV or blue light, and thus has significant potential for use in monitoring gene expression in living cells.

Other particular examples are the enzyme chloramphenicol acetyltransferase (CAT) which may be employed with a radiolabelled substrate, firefly and bacterial luciferase, and the bacterial enzymes β -galactosidase and β -glucuronidase. Other marker genes within this class are well known to those of skill in the art, and are suitable for use in the present invention.

2. Selection

Another class of reporter genes which confer detectable characteristics on a host cell are those which encode polypeptides, generally enzymes, which render their transformants resistant against toxins. Examples of this class of reporter genes are the *neo* gene (Colberre-Garapin *et al.*, 1981) which protects host cells against toxic levels of the antibiotic G418, the gene conferring streptomycin resistance (U. S. Patent 4,430,434), the gene conferring hygromycin B resistance (Santerre *et al.*, 1984; U. S. Patents 4,727,028, 4,960,704 and 4,559,302), a gene encoding dihydrofolate reductase, which confers resistance to methotrexate (Alt *et al.*, 1978), the enzyme HPRT, along with many others well known in the art (Kaufman, 1990).

E. Biological Functional Equivalents

While the present invention contemplates the use of tumor suppressor proteins, exemplified by the retinoblastoma protein, which contain modifications within the N-terminal region which confer equal or greater tumor suppression activity on the resultant protein, alteration of the unmodified C-terminal portion of the protein such that biological activity is maintained also falls within the scope of the present invention.

As mentioned above, modification and changes may be made in the structure of, for example, the retinoblastoma protein, and still obtain a molecule having like or otherwise desirable characteristics. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of tumor suppression activity. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence (or, of course, its underlying DNA coding sequence) and nevertheless obtain a protein with like (agonistic) properties. Equally, the same considerations may be employed to create a protein or polypeptide with countervailing (*e.g.*, antagonistic) properties. It is thus contemplated by the inventors that various changes may be made in the sequence of tumor suppressor proteins or peptides (or underlying DNA) without appreciable loss of their biological utility or activity.

In terms of functional equivalents, It is also well understood by the skilled artisan that, inherent in the definition of a biologically functional equivalent protein or peptide, is the concept that there is a limit to the number of changes that may be made within a defined portion of the molecule and still result in a molecule with an acceptable level of equivalent biological activity. Biologically functional equivalent peptides are thus defined herein as those peptides in which certain, not most or all, of the amino acids may be substituted. Of course, a plurality of distinct proteins/peptides with different substitutions may easily be made and used in accordance with the invention.

It is also well understood that where certain residues are shown to be particularly important to the biological or structural properties of a protein or peptide, *e.g.*, residues in active sites, such residues may not generally be exchanged.

Conservative substitutions well known in the art include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine; isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to

arginine, glutamine, or glutamate; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine.

5 In making such changes, the hydropathic index of amino acids may be considered. Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics, these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate
10 (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte & Doolittle, 1982, incorporated herein by reference). It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indices are within ± 2 is preferred, those which are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, i.e. with a biological property of the protein. use this shorter portion for non-immunological
25 stuff It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been
30 assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate ($+3.0 \pm 1$); glutamate

(+3.0 \pm 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 \pm 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

5 In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those which are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

10 While discussion has focused on functionally equivalent polypeptides arising from amino acid changes, it will be appreciated that these changes may be effected by alteration of the encoding DNA; taking into consideration also that the genetic code is degenerate and that two or more codons may code for the same amino acid. Two tables of amino acids and their codons is presented below for use in such embodiments, as well as for other uses, such as in the design of probes and primers and the like.

Table 1 - Preferred Human DNA Codons

<u>Amino Acids</u>		<u>Codons</u>							
Alanine	Ala	A	GCC	GCT	GCA	GCG			
Cysteine	Cys	C	TGC	TGT					
Aspartic acid	Asp	D	GAC	GAT					
Glutamic acid	Glu	E	GAG	GAA					
Phenylalanine	Phe	F	TTC	TTT					
Glycine	Gly	G	GGC	GGG	GGA	GGT			
Histidine	His	H	CAC	CAT					
Isoleucine	Ile	I	ATC	ATT	ATA				
Lysine	Lys	K	AAG	AAA					
Leucine	Leu	L	CTG	CTC	TTG	CTT	CTA	TTA	
Methionine	Met	M	ATG						
Asparagine	Asn	N	AAC	AAT					
Proline	Pro	P	CCC	CCT	CCA	CCG			
Glutamine	Gln	Q	CAG	CAA					
Arginine	Arg	R	CGC	AGG	CGG	AGA	CGA	<u>CGT</u>	
Serine	Ser	S	AGC	TCC	TCT	AGT	TCA	<u>TCG</u>	
Threonine	Thr	T	ACC	ACA	ACT	ACG			
Valine	Val	V	GTG	GTC	GTT	GTA			
Tryptophan	Trp	W	TGG						
Tyrosine	Tyr	Y	TAC	TAT					

Codon prevalence shown as decreasing from left (most prevalent) to right (least prevalent).

Underlined codons are those used less than 5 times per one thousand codons.

Table 2 - Preferred Human RNA Codons

<u>Amino Acids</u>			<u>Codons</u>						
Alanine	Ala	A	GCC	GCU	GCA	GCG			
Cysteine	Cys	C	UGC	UGU					
Aspartic acid	Asp	D	GAC	GAU					
Glutamic acid	Glu	E	GAG	GAA					
Phenylalanine	Phe	F	UUC	UUU					
Glycine	Gly	G	GGC	GGG	GGA	GGU			
Histidine	His	H	CAC	CAU					
Isoleucine	Ile	I	AUC	AUU	AUA				
Lysine	Lys	K	AAG	AAA					
Leucine	Leu	L	CUG	CUC	UUG	CUU	CUA	UUA	
Methionine	Met	M	AUG						
Asparagine	Asn	N	AAC	AAU					
Proline	Pro	P	CCC	CCU	CCA	CCG			
Glutamine	Gln	Q	CAG	CAA					
Arginine	Arg	R	CGC	AGG	CGG	AGA	CGA	<u>CGU</u>	
Serine	Ser	S	AGC	UCC	UCU	AGU	UCA	<u>UCG</u>	
Threonine	Thr	T	ACC	ACA	ACU	ACG			
Valine	Val	V	GUG	GUC	GUU	GUA			
Tryptophan	Trp	W	UGG						
Tyrosine	Tyr	Y	UAC	UAU					

Codon prevalence shown as decreasing from left (most prevalent) to right (least prevalent).

Underlined codons are those used less than 5 times per one thousand codons.

F. Mutagenesis

Mutagenesis may be performed in accordance with any of the techniques known in the art such as and not limited to synthesizing an oligonucleotide having one or more mutations within the sequence of a particular tumor suppressor or cytokine protein. In particular, site-specific

mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent proteins or peptides, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA.

Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to about 75 nucleotides or more in length is preferred, with about 10 to about 25 or more residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed mutagenesis which eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double stranded vector which includes within its sequence a DNA sequence which encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex

vector is then used to transform or transfect appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement. A genetic selection scheme was devised by Kunkel *et al.* (1987) to enrich for clones incorporating the mutagenic oligonucleotide.

5

10

Alternatively, the use of PCRTM with commercially available thermostable enzymes such as *Taq* polymerase may be used to incorporate a mutagenic oligonucleotide primer into an amplified DNA fragment that can then be cloned into an appropriate cloning or expression vector. The PCRTM-mediated mutagenesis procedures of Tomic *et al.* (1990) and Upender *et al.* (1995) provide two examples of such protocols. A PCRTM employing a thermostable ligase in addition to a thermostable polymerase may also be used to incorporate a phosphorylated mutagenic oligonucleotide into an amplified DNA fragment that may then be cloned into an appropriate cloning or expression vector. The mutagenesis procedure described by Michael (1994) provides an example of one such protocol.

15

20

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

25

30

As used herein, the term “oligonucleotide directed mutagenesis procedure” refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term “oligonucleotide directed mutagenesis procedure” is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of

complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U.S. Patent 4,237,224, specifically incorporated herein by reference in its entirety.

G. Pharmaceutically Acceptable Compositions and Routes of Administration

Where clinical applications are contemplated, it will be necessary to prepare pharmaceutical compositions of the proteins, nucleic acids, including vectors such as tetracycline-regulated vectors, recombinant viruses and cells in a form appropriate for the intended application. Generally, this will entail preparing compositions that are essentially free of pyrogens, as well as other impurities that could be harmful to humans or animals.

One will generally desire to employ appropriate salts and buffers to render the compositions suitable for introduction into a patient. Aqueous compositions of the present invention comprise an effective amount of the therapeutic agent dissolved or dispersed in a pharmaceutically acceptable carrier or aqueous medium, and preferably encapsulated. The phrase "pharmaceutically or pharmacologically acceptable" refer to molecular entities and compositions that do not produce adverse, allergic, or other untoward reactions when administered to an animal or a human. As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutically active substances is well know in the art. Except insofar as any conventional media or agent is incompatible with the vectors or cells of the present invention, its use in therapeutic compositions is contemplated. Supplementary active ingredients, such as other anti-cancer agents, can also be incorporated into the compositions.

Solutions of the active ingredients as free base or pharmacologically acceptable salts can be prepared in water suitably mixed with surfactant, such as hydroxypropylcellulose. Dispersions also can be prepared in glycerol, liquid polyethylene glycols, mixtures thereof and in

oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent growth of microorganisms. Intravenous vehicles include fluid and nutrient replenishers. Preservatives include antimicrobial agents, anti-oxidants, chelating agents and inert gases. The pH and exact concentration of the various components in the pharmaceutical are adjusted according to well-known parameters.

An effective amount of the viruses or cells is determined based on the intended goal. The term "unit dose" refers to a physically discrete unit suitable for use in a subject, each unit containing a predetermined quantity of the therapeutic composition calculated to produce the desired response in association with its administration, *i.e.*, the appropriate route and treatment regimen. The quantity to be administered, both according to number of treatments and unit dose, depends on the subject to be treated, the state of the subject, and the protection desired. Precise amounts of the therapeutic composition also depend on the judgment of the practitioner and are peculiar to each individual.

1. Parenteral Administration

The active compositions of the present invention will often be formulated for parenteral administration, *e.g.*, formulated for injection *via* the intravenous, intramuscular, sub-cutaneous, intratumoral, peritumoral or even intraperitoneal routes. The preparation of an aqueous composition that contains a second agent(s) as active ingredients will be known to those of skill in the art in light of the present disclosure. Typically, such compositions can be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for using to prepare solutions or suspensions upon the addition of a liquid prior to injection can also be prepared; and the preparations can also be emulsified.

Solutions of the active compounds as free base or pharmacologically acceptable salts can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions; formulations including sesame oil, peanut oil or aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions.

5 In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi.

10 The active compounds may be formulated into a composition in a neutral or salt form. Pharmaceutically acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

20 The carrier can also be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged
25 absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

30 Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by

incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the particular methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous, intratumoral, peritumoral and intraperitoneal administration. In this connection, sterile aqueous media which can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage could be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject.

2. Other Routes of Administration

In addition to the compounds formulated for parenteral administration, such as intravenous or intramuscular injection, other pharmaceutically acceptable forms include, *e.g.*, tablets or other solids for oral administration; time release capsules; and any other form currently used, including cremes, lotions, mouthwashes, inhalants and the like.

The expression vectors and delivery vehicles of the present invention may include classic pharmaceutical preparations. Administration of these compositions according to the present invention will be *via* any common route so long as the target tissue is available *via* that route. This includes oral, nasal, buccal, rectal, vaginal or topical. Alternatively, administration may be by orthotopic, intradermal, subcutaneous, intramuscular, intraperitoneal or intravenous injection.

The injection can be general, regional, local or direct injection, for example, of a tumor. Also contemplated is injection of a resected tumor bed, and continuous perfusion *via* catheter. Such compositions would normally be administered as pharmaceutically acceptable compositions, described *supra*.

5

10 The vectors of the present invention are advantageously administered in the form of injectable compositions either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection also may be prepared. These preparations also may be emulsified. A typical compositions for such purposes comprises a 50 mg or up to about 100 mg of human serum albumin per milliliter of phosphate buffered saline. Other pharmaceutically acceptable carriers include aqueous solutions, non-toxic excipients, including salts, preservatives, buffers and the like. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oil and injectable organic esters, such as theylolate. Aqueous carriers include water, alcoholic/aqueous solutions, saline solutions, parenteral vehicles such as sodium chloride, Ringer's dextrose, *etc.* Intravenous vehicles include fluid and nutrient replenishers. Preservatives include antimicrobial agents, anti-oxidants, chelating agents and inert gases. The pH and exact concentration of the various components in the pharmaceutical are adjusted according to well known parameters.

20 Additional formulations are suitable for oral administration. Oral formulations include such typical excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate and the like. The compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders. When the route is topical, the form may be a cream, ointment, salve or spray.

25 An effective amount of the therapeutic agent is determined based on the intended goal. The term "unit dose" refers to a physically discrete unit suitable for use in a subject, each unit containing a predetermined quantity of the therapeutic composition calculated to produce the desired response in association with its administration, *i.e.*, the appropriate route and treatment

30

regimen. The quantity to be administered, both according to number of treatments and unit dose, depends on the subject to be treated, the state of the subject and the protection desired. Precise amounts of the therapeutic composition also depend on the judgment of the practitioner and are peculiar to each individual.

5

In certain cases, the therapeutic formulations of the invention could also be prepared in forms suitable for topical administration, such as in cremes and lotions. These forms may be used for treating skin-associated diseases, such as various sarcomas.

10

Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms, such as the type of injectable solutions described above, with even drug release capsules and the like being employable.

15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995

H. Chemotherapeutic Agents

The methods of the present invention may be combined with any other methods generally employed in the treatment of the particular disease or disorder that the patient exhibits. For example, in connection with the treatment of solid tumors, the methods of the present invention may be used in combination with classical approaches, such as surgery, radiotherapy and the like. So long as a particular therapeutic approach is not known to be detrimental in itself, or counteracts the effectiveness of the tumor suppressor therapy, its combination with the present invention is contemplated. When one or more agents are used in combination with cytokine gene therapy and/or tumor suppressor gene therapy, there is no requirement for the combined results to be additive of the effects observed when each treatment is conducted separately, although this is evidently desirable, and there is no particular requirement for the combined treatment to exhibit synergistic effects, although this is certainly possible and advantageous.

25

30

In terms of surgery, any surgical intervention may be practiced in combination with the present invention. In connection with radiotherapy, any mechanism for inducing DNA damage locally within tumor cells is contemplated, such as γ -irradiation, X-rays, UV-irradiation,

microwaves and even electronic emissions and the like. The directed delivery of radioisotopes to tumor cells is also contemplated, and this may be used in connection with a targeting antibody or other targeting means. Cytokine therapy also has proven to be an effective partner for combined therapeutic regimens. Various cytokines may be employed in such combined approaches.

5 Examples of cytokines include IL-1 α IL-1 β , IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TGF- β , GM-CSF, M-CSF, G-CSF, TNF α , TNF β , LAF, TCGF, BCGF, TRF, BAF, BDG, MP, LIF, OSM, TMF, PDGF, IFN- α , IFN- β , IFN- γ . Cytokines are administered according to standard regimens, consistent with clinical indications such as the condition of the patient and relative toxicity of the cytokine. Below is an exemplary, but in no

10 way limiting, table of cytokine genes contemplated for use in certain embodiments of the present invention.

Table 3

Cytokine	Reference
human IL-1 α	March <i>et al.</i> , <i>Nature</i> , 315:641, 1985
murine IL-1 α	Lomedico <i>et al.</i> , <i>Nature</i> , 312:458, 1984
human IL-1 β	March <i>et al.</i> , <i>Nature</i> , 315:641, 1985; Auron <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 81:7907, 1984
murine IL-1 β	Gray, <i>J. Immunol.</i> , 137:3644, 1986; Telford, <i>Nucl. Acids Res.</i> , 14:9955, 1986
human IL-1ra	Eisenberg <i>et al.</i> , <i>Nature</i> , 343:341, 1990
human IL-2	Taniguchi <i>et al.</i> , <i>Nature</i> , 302:305, 1983; Maeda <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i> , 115:1040, 1983
human IL-2	Taniguchi <i>et al.</i> , <i>Nature</i> , 302:305, 1983
human IL-3	Yang <i>et al.</i> , <i>Cell</i> , 47:3, 1986
murine IL-3	Yokota <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 81:1070, 1984; Fung <i>et al.</i> , <i>Nature</i> , 307:233, 1984; Miyatake <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 82:316, 1985
human IL-4	Yokota <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 83:5894, 1986
murine IL-4	Norma <i>et al.</i> , <i>Nature</i> , 319:640, 1986; Lee <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 83:2061, 1986
human IL-5	Azuma <i>et al.</i> , <i>Nucl. Acids Res.</i> , 14:9149, 1986
murine IL-5	Kinashi <i>et al.</i> , <i>Nature</i> , 324:70, 1986; Mizuta <i>et al.</i> , <i>Growth Factors</i> , 1:51, 1988
human IL-6	Hirano <i>et al.</i> , <i>Nature</i> , 324:73, 1986
murine IL-6	Van Snick <i>et al.</i> , <i>Eur. J. Immunol.</i> , 18:193, 1988
human IL-7	Goodwin <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 86:302, 1989

Cytokine	Reference
murine IL-7	Namen <i>et al.</i> , <i>Nature</i> , 333:571, 1988
human IL-8	Schmid <i>et al.</i> , <i>J. Immunol.</i> , 139:250, 1987; Matsushima <i>et al.</i> , <i>J. Exp. Med.</i> , 167:1883, 1988; Lindley <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 85:9199, 1988
human IL-9	Renauld <i>et al.</i> , <i>J. Immunol.</i> , 144:4235, 1990
murine IL-9	Renauld <i>et al.</i> , <i>J. Immunol.</i> , 144:4235, 1990
human Angiogenin	Kurachi <i>et al.</i> , <i>Biochemistry</i> , 24:5494, 1985
human GRO α	Richmond <i>et al.</i> , <i>EMBO J.</i> , 7:2025, 1988
murine MIP-1 α	Davatellis <i>et al.</i> , <i>J. Exp. Med.</i> , 167:1939, 1988
murine MIP-1 β	Sherry <i>et al.</i> , <i>J. Exp. Med.</i> , 168:2251, 1988
human MIF	Weiser <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 86:7522, 1989
human G-CSF	Nagata <i>et al.</i> , <i>Nature</i> , 319:415, 1986; Souza <i>et al.</i> , <i>Science</i> , 232:61, 1986
human GM-CSF	Cantrell <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 82:6250, 1985; Lee <i>et al.</i> , <i>Proc. Natl. Acad. Sci. USA</i> , 82:4360, 1985; Wong <i>et al.</i> , <i>Science</i> , 228:810, 1985
murine GM-CSF	Gough <i>et al.</i> , <i>EMBO J.</i> , 4:645, 1985
human M-CSF	Wong, <i>Science</i> , 235:1504, 1987; Kawasaki, <i>Science</i> , 230:291, 1985; Ladner, <i>EMBO J.</i> , 6:2693, 1987
human EGF	Smith <i>et al.</i> , <i>Nucl. Acids Res.</i> , 10:4467, 1982; Bell <i>et al.</i> , <i>Nucl. Acids Res.</i> , 14:8427, 1986
human TGF- α	Derynck <i>et al.</i> , <i>Cell</i> , 38:287, 1984
human FGF acidic	Jaye <i>et al.</i> , <i>Science</i> , 233:541, 1986; Gimenez-Gallego <i>et al.</i> , <i>Biochem. Biophys. Res. Commun.</i> , 138:611, 1986; Harper <i>et al.</i> , <i>Biochem.</i> , 25:4097, 1986
human β -ECGF	Jaye <i>et al.</i> , <i>Science</i> , 233:541, 1986
human FGF basic	Abraham <i>et al.</i> , <i>EMBO J.</i> , 5:2523, 1986; Sommer <i>et al.</i> , <i>Biochem. Biophys. Res. Comm.</i> , 144:543, 1987
murine IFN- β	Higashi <i>et al.</i> , <i>J. Biol. Chem.</i> , 258:9522, 1983; Kuga, <i>Nucl. Acids Res.</i> , 17:3291, 1989
human IFN- γ	Gray <i>et al.</i> , <i>Nature</i> , 295:503, 1982; Devos <i>et al.</i> , <i>Nucl. Acids Res.</i> , 10:2487, 1982; Rinderknecht, <i>J. Biol. Chem.</i> , 259:6790, 1984
human IGF-I	Jansen <i>et al.</i> , <i>Nature</i> , 306:609, 1983; Rotwein <i>et al.</i> , <i>J. Biol. Chem.</i> , 261:4828, 1986
human IGF-II	Bell <i>et al.</i> , <i>Nature</i> , 310:775, 1984
human β -NGF chain	Ullrich <i>et al.</i> , <i>Nature</i> , 303:821, 1983
human PDGF A chain	Betsholtz <i>et al.</i> , <i>Nature</i> , 320:695, 1986
human PDGF B chain	Johnsson <i>et al.</i> , <i>EMBO J.</i> , 3:921, 1984; Collins <i>et al.</i> , <i>Nature</i> , 316:748, 1985
human TGF- β 1	Derynck <i>et al.</i> , <i>Nature</i> , 316:701, 1985
human TNF- α	Pennica <i>et al.</i> , <i>Nature</i> , 312:724, 1984; Fransen <i>et al.</i> , <i>Nucl. Acids Res.</i> , 13:4417, 1985

Cytokine	Reference
human TNF- β	Gray <i>et al.</i> , <i>Nature</i> , 312:721, 1984
murine TNF- β	Gray <i>et al.</i> , <i>Nucl. Acids Res.</i> , 15:3937, 1987

Compositions of the present invention can have an effective amount of an engineered virus or cell for therapeutic administration in combination with an effective amount of a compound (second agent) that is a chemotherapeutic agent as exemplified below. Such compositions will generally be dissolved or dispersed in a pharmaceutically acceptable carrier or aqueous medium. A wide variety of chemotherapeutic agents may be used in combination with the therapeutic genes of the present invention. These can be, for example, agents that directly cross-link DNA, agents that intercalate into DNA, and agents that lead to chromosomal and mitotic aberrations by affecting nucleic acid synthesis.

Irrespective of the mechanisms by which the enhanced tumor destruction is achieved, the combined treatment aspects of the present invention have evident utility in the effective treatment of disease. To use the compositions of the present invention in combination with the administration of a chemotherapeutic agent, one would simply administer to an animal at least a first modified retinoblastoma tumor suppressor as disclosed herein in combination with the chemotherapeutic agent in a manner effective to result in their combined anti-tumor actions within the animal. These agents would therefore be provided in an amount effective and for a period of time effective to result in their combined presence and their combined actions in the tumor environment. To achieve this goal, the modified retinoblastoma tumor suppressor and chemotherapeutic agents may be administered to the animal simultaneously, either in a single composition or as two distinct compositions using different administration routes.

Alternatively, the modified retinoblastoma tumor suppressor treatment may precede or follow the chemotherapeutic agent treatment by intervals ranging from minutes to weeks. In embodiments where the chemotherapeutic factor and modified retinoblastoma tumor suppressor are applied separately to the animal, one would generally ensure that a significant period of time did not expire between the time of each delivery, such that the chemotherapeutic agent and modified retinoblastoma tumor suppressor composition would still be able to exert an

advantageously combined effect on the tumor. In such instances, it is contemplated that one would contact the tumor with both agents within about 5 minutes to about one week of each other and, more preferably, within about 12-72 hours of each other, with a delay time of only about 12-48 hours being most preferred. In some situations, it may be desirable to extend the time period for treatment significantly, where several days (2, 3, 4, 5, 6 or 7) or even several weeks (1, 2, 3, 4, 5, 6, 7 or 8) lapse between the respective administrations. It also is conceivable that more than one administrations of either the modified retinoblastoma tumor suppressor or the chemotherapeutic agent will be desired. To achieve tumor regression, both agents are delivered in a combined amount effective to inhibit its growth, irrespective of the times for administration.

A variety of chemotherapeutic agents are intended to be of use in the combined treatment methods disclosed herein. Chemotherapeutic agents contemplated as exemplary include, *e.g.*, etoposide (VP-16), adriamycin, 5-fluorouracil (5FU), camptothecin, actinomycin-D, mitomycin C, cisplatin (CDDP) and even hydrogen peroxide.

As will be understood by those of ordinary skill in the art, the appropriate doses of chemotherapeutic agents will be generally around those already employed in clinical therapies wherein the chemotherapeutics are administered alone or in combination with other chemotherapeutics. By way of example only, agents such as cisplatin, and other DNA alkylating may be used. Cisplatin has been widely used to treat cancer, with efficacious doses used in clinical applications of 20 mg/m² for 5 days every three weeks for a total of three courses. Cisplatin is not absorbed orally and must therefore be delivered *via* injection intravenously, subcutaneously, intratumorally or intraperitoneally.

Agents that directly cross-link nucleic acids, specifically DNA, are envisaged and are shown herein, to eventuate DNA damage leading to a synergistic antineoplastic combination. Agents such as cisplatin, and other DNA alkylating agents may be used.

Further useful agents include compounds that interfere with DNA replication, mitosis and chromosomal segregation. Such chemotherapeutic compounds include adriamycin, also known

as doxorubicin, etoposide, verapamil, podophyllotoxin, and the like. Widely used in a clinical setting for the treatment of neoplasms, these compounds are administered through bolus injections intravenously at doses ranging from 25-75 mg/m² at 21 day intervals for adriamycin, to 35-50 mg/m² for etoposide intravenously or double the intravenous dose orally.

5

Agents that disrupt the synthesis and fidelity of polynucleotide precursors may also be used. Particularly useful are agents that have undergone extensive testing and are readily available. As such, agents such as 5-fluorouracil (5-FU) are preferentially used by neoplastic tissue, making this agent particularly useful for targeting to neoplastic cells. Although quite toxic, 5-FU, is applicable in a wide range of carriers, including topical, however intravenous administration with doses ranging from 3 to 15 mg/kg/day being commonly used.

10

15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995

Plant alkaloids such as taxol are also contemplated for use in certain aspects of the present invention. Taxol is an experimental antimitotic agent, isolated from the bark of the ash tree, *Taxus brevifolia*. It binds to tubulin (at a site distinct from that used by the vinca alkaloids) and promotes the assembly of microtubules. Taxol is currently being evaluated clinically; it has activity against malignant melanoma and carcinoma of the ovary. Maximal doses are 30 mg/m² per day for 5 days or 210 to 250 mg/m² given once every 3 weeks. Of course, all of these dosages are exemplary, and any dosage in-between these points is also expected to be of use in the invention.

20

25

Exemplary chemotherapeutic agents that are useful in connection with combined therapy are listed in Table 4. Each of the agents listed therein are exemplary and by no means limiting. The skilled artisan is directed to "Remington's Pharmaceutical Sciences" 15th Edition, chapter 33, in particular pages 624-652. Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

30

Table 4

Chemotherapeutic Agents Useful In Neoplastic Disease

Class	Type Of Agent	Nonproprietary Names (Other Names)	Disease
<i>Alkylating Agents</i>	Nitrogen Mustards	Mechlorethamine (HN ₂)	Hodgkin's disease, non-Hodgkin's lymphomas
		Cyclophosphamide Ifosfamide	Acute and chronic lymphocytic leukemias, Hodgkin's disease, non-Hodgkin's lymphomas, multiple myeloma, neuroblastoma, breast, ovary, lung, Wilms' tumor, cervix, testis, soft-tissue sarcomas
		Melphalan (L-sarcolysin)	Multiple myeloma, breast, ovary
		Chlorambucil	Chronic lymphocytic leukemia, primary macroglobulinemia, Hodgkin's disease, non- Hodgkin's lymphomas
	Ethylenimenes and Methylmelamines	Hexamethylmelamine	Ovary
		Thiotepa	Bladder, breast, ovary
	Alkyl Sulfonates	Busulfan	Chronic granulocytic leukemia
	Nitrosoureas	Carmustine (BCNU)	Hodgkin's disease, non-Hodgkin's lymphomas, primary brain tumors, multiple myeloma, malignant melanoma
		Lomustine (CCNU)	Hodgkin's disease, non-Hodgkin's lymphomas, primary brain tumors, small-cell lung
		Semustine (methyl-CCNU)	Primary brain tumors, stomach, colon
		Streptozocin (streptozotocin)	Malignant pancreatic insulinoma, malignant carcinoid
	Triazines	Dacarbazine (DTIC; dimethyltriazenoimidaz olecarboxamide)	Malignant melanoma, Hodgkin's disease, soft- tissue sarcomas
<i>Antimetabolites</i>	Folic Acid Analogs	Methotrexate (amethopterin)	Acute lymphocytic leukemia, choriocarcinoma, mycosis fungoides, breast, head and neck, lung, osteogenic sarcoma
	Pyrimidine Analogs	Fluorouracil (5-fluorouracil; 5-FU)	Breast, colon, stomach, pancreas, ovary, head and neck, urinary bladder, premalignant skin lesions (topical)
		Floxuridine (fluorode- oxyuridine; FdR)	
		Cytarabine (cytosine arabioside)	Acute granulocytic and acute lymphocytic leukemias
	Purine Analogs and Related Inhibitors	Mercaptopurine (6-mercaptopurine; 6-MP)	Acute lymphocytic, acute granulocytic and chronic granulocytic leukemias
		Thioguanine (6-thioguanine; TG)	Acute granulocytic, acute lymphocytic and chronic granulocytic leukemias

Class	Type Of Agent	Nonproprietary Names (Other Names)	Disease
		Pentostatin (2-deoxycoformycin)	Hairy cell leukemia, mycosis fungoides, chronic lymphocytic leukemia
<i>Natural Products</i>	Vinca Alkaloids	Vinblastine (VLB)	Hodgkin's disease, non-Hodgkin's lymphomas, breast, testis
		Vincristine	Acute lymphocytic leukemia, neuroblastoma, Wilms' tumor, rhabdomyosarcoma, Hodgkin's disease, non-Hodgkin's lymphomas, small-cell lung
	Epipodophyllotoxins	Etoposide (VP16) Tertiposide	Testis, small-cell lung and other lung, breast, Hodgkin's disease, non-Hodgkin's lymphomas, acute granulocytic leukemia, Kaposi's sarcoma
	Antibiotics	Dactinomycin (actinomycin D)	Choriocarcinoma, Wilms' tumor, rhabdomyosarcoma, testis, Kaposi's sarcoma
		Daunorubicin (daunomycin; rubidomycin)	Acute granulocytic and acute lymphocytic leukemias
		Doxorubicin	Soft-tissue, osteogenic and other sarcomas; Hodgkin's disease, non-Hodgkin's lymphomas, acute leukemias, breast, genitourinary, thyroid, lung, stomach, neuroblastoma
		Bleomycin	Testis, head and neck, skin, esophagus, lung and genitourinary tract; Hodgkin's disease, non-Hodgkin's lymphomas
	Antibiotics, continued	Plicamycin (mithramycin)	Testis, malignant hypercalcemia
		Mitomycin (mitomycin C)	Stomach, cervix, colon, breast, pancreas, bladder, head and neck
	Enzymes	L-Asparaginase	Acute lymphocytic leukemia
	Biological Response Modifiers	Interferon alfa	Hairy cell leukemia, Kaposi's sarcoma, melanoma, carcinoid, renal cell, ovary, bladder, non-Hodgkin's lymphomas, mycosis fungoides, multiple myeloma, chronic granulocytic leukemia
<i>Miscellaneous Agents</i>	Platinum Coordination Complexes	Cisplatin (<i>cis</i> -DDP) Carboplatin	Testis, ovary, bladder, head and neck, lung, thyroid, cervix, endometrium, neuroblastoma, osteogenic sarcoma
	Anthracenedione	Mitoxantrone	Acute granulocytic leukemia, breast
	Substituted Urea	Hydroxyurea	Chronic granulocytic leukemia, polycythemia vera, essential thrombocytosis, malignant melanoma
	Methyl Hydrazine Derivative	Procarbazine (N-methylhydrazine, MIH)	Hodgkin's disease
	Adrenocortical Suppressant	Mitotane (<i>o,p'</i> -DDD)	Adrenal cortex
		Aminoglutethimide	Breast

Class	Type Of Agent	Nonproprietary Names (Other Names)	Disease
<i>Hormones and Antagonists</i>	Adrenocorticosteroids	Prednisone (several other equivalent preparations available)	Acute and chronic lymphocytic leukemias, non- Hodgkin's lymphomas, Hodgkin's disease, breast
	Progestins	Hydroxyprogesterone caproate Medroxyprogesterone acetate Megestrol acetate	Endometrium, breast
	Estrogens	Diethylstilbestrol Ethinyl estradiol (other preparations available)	Breast, prostate
	Antiestrogen	Tamoxifen	Breast
	Androgens	Testosterone propionate Fluoxymesterone (other preparations available)	Breast
	Antiandrogen	Flutamide	Prostate
	Gonadotropin-releasing hormone analog	Leuprolide	Prostate

I. Protein Purification

Certain aspects of the present invention concern the purification, and in particular embodiments, the substantial purification, of an encoded protein or peptide. The term "purified protein or peptide " as used herein, is intended to refer to a composition, isolatable from other components, wherein the protein or peptide is purified to any degree relative to its naturally-obtainable state. A purified protein or peptide therefore also refers to a protein or peptide, free from the environment in which it may naturally occur.

Generally, "purified" will refer to a protein or peptide composition that has been subjected to fractionation to remove various other components, and which composition substantially retains its expressed biological activity. Where the term "substantially purified" is used, this designation will refer to a composition in which the protein or peptide forms the major component of the composition, such as constituting about 50% or more of the proteins in the composition.

Various methods for quantifying the degree of purification of the protein or peptide will be known to those of skill in the art in light of the present disclosure. These include, for example, determining the specific activity of an active fraction, or assessing the number of polypeptides within a fraction by SDS/PAGE analysis. A preferred method for assessing the purity of a fraction is to calculate the specific activity of the fraction, to compare it to the specific activity of the initial extract, and to thus calculate the degree of purity, herein assessed by a "-fold purification number". The actual units used to represent the amount of activity will, of course, be dependent upon the particular assay technique chosen to follow the purification and whether or not the expressed protein or peptide exhibits a detectable activity.

Various techniques suitable for use in protein purification will be well known to those of skill in the art. These include, for example, precipitation with ammonium sulphate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite and affinity chromatography; isoelectric focusing; gel electrophoresis; and combinations of such and other techniques. As is generally known in the art, it is believed that the order of conducting the various purification steps may be changed, or that certain steps may be omitted, and still result in a suitable method for the preparation of a substantially purified protein or peptide.

There is no general requirement that the protein or peptide always be provided in their most purified state. Indeed, it is contemplated that less substantially purified products will have utility in certain embodiments. Partial purification may be accomplished by using fewer purification steps in combination, or by utilizing different forms of the same general purification scheme. For example, it is appreciated that a cation-exchange column chromatography performed utilizing an HPLC apparatus will generally result in a greater -fold purification than the same technique utilizing a low pressure chromatography system. Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein.

It is known that the migration of a polypeptide can vary, sometimes significantly, with different conditions of SDS/PAGE (Capaldi *et al.*, 1977). It will therefore be appreciated that under differing electrophoresis conditions, the apparent molecular weights of purified or partially purified expression products may vary.

5

High Performance Liquid Chromatography (HPLC) is characterized by a very rapid separation with extraordinary resolution of peaks. This is achieved by the use of very fine particles and high pressure to maintain an adequate flow rate. Separation can be accomplished in a matter of minutes, or at most an hour. Moreover, only a very small volume of the sample is needed because the particles are so small and close-packed that the void volume is a very small fraction of the bed volume. Also, the concentration of the sample need not be very great because the bands are so narrow that there is very little dilution of the sample.

10

Gel chromatography, or molecular sieve chromatography, is a special type of partition chromatography that is based on molecular size. The theory behind gel chromatography is that the column, which is prepared with tiny particles of an inert substance that contain small pores, separates larger molecules from smaller molecules as they pass through or around the pores, depending on their size. As long as the material of which the particles are made does not adsorb the molecules, the sole factor determining rate of flow is the size. Hence, molecules are eluted from the column in decreasing size, so long as the shape is relatively constant. Gel chromatography is unsurpassed for separating molecules of different size because separation is independent of all other factors such as pH, ionic strength, temperature, *etc.* There also is virtually no adsorption, less zone spreading and the elution volume is related in a simple manner to molecular weight.

20

25

Affinity chromatography is a chromatographic procedure that relies on the specific affinity between a substance to be isolated and a molecule that it can specifically bind to. This is a receptor-ligand type interaction. The column material is synthesized by covalently coupling one of the binding partners to an insoluble matrix. The column material is then able to

specifically adsorb the substance from the solution. Elution occurs by changing the conditions to those in which binding will not occur (alter pH, ionic strength, temperature, *etc.*).

A particular type of affinity chromatography useful in the purification of carbohydrate containing compounds is lectin affinity chromatography. Lectins are a class of substances that bind to a variety of polysaccharides and glycoproteins. Lectins are usually coupled to agarose by cyanogen bromide. Concanavalin A coupled to Sepharose was the first material of this sort to be used and has been widely used in the isolation of polysaccharides and glycoproteins other lectins that have been include lentil lectin, wheat germ agglutinin which has been useful in the purification of N-acetyl glucosaminyl residues and *Helix pomatia* lectin. Lectins themselves are purified using affinity chromatography with carbohydrate ligands. Lactose has been used to purify lectins from castor bean and peanuts; maltose has been useful in extracting lectins from lentils and jack bean; N-acetyl-D galactosamine is used for purifying lectins from soybean; N-acetyl glucosaminyl binds to lectins from wheat germ; D-galactosamine has been used in obtaining lectins from clams and L-fucose will bind to lectins from lotus.

The matrix should be a substance that itself does not adsorb molecules to any significant extent and that has a broad range of chemical, physical and thermal stability. The ligand should be coupled in such a way as to not affect its binding properties. The ligand should also provide relatively tight binding. And it should be possible to elute the substance without destroying the sample or the ligand. One of the most common forms of affinity chromatography is immunoaffinity chromatography.

L. Use of Cells in Bioreactors

The ability to produce biologically active polypeptides is increasingly important to the pharmaceutical industry. The present invention discloses compositions and methods for the efficient regulated expression of, for example, tumor suppressor genes in cells, allowing for the production of these proteins *in vitro* from previously refractory cell types.

Over the last decade, advances in biotechnology have led to the production of important proteins and factors from bacteria, yeast, insect cells and from mammalian cell culture. Mammalian cultures have advantages over cultures derived from the less advanced lifeforms in their ability to post-translationally process complex protein structures such as disulfide-dependent folding and glycosylation. Indeed, mammalian cell culture is now the preferred source of a number of important proteins for use in human and animal medicine, especially those which are relatively large, complex or glycosylated.

Development of mammalian cell culture for production of pharmaceuticals has been greatly aided by the development in molecular biology of techniques for design and construction of vector systems highly efficient in mammalian cell cultures, a battery of useful selection markers, gene amplification schemes and a more comprehensive understanding of the biochemical and cellular mechanisms involved in procuring the final biologically-active molecule from the introduced vector.

However, the traditional selection of cell types for expressing heterologous proteins has generally been limited to the more "common" cell types such as CHO cells, BHK cells, C127 cells and myeloma cells. In many cases, these cell types were selected because there was a great deal of preexisting literature on the cell type or the cell was simply being carried in the laboratory at the time the effort was made to express a peptide product. Frequently, factors which affect the downstream (*e.g.*, beyond the T-75 flask) side of manufacturing scale-up were not considered before selecting the cell line as the host for the expression system.

Aspects of the present invention take advantage of the biochemical and cellular capacities of mammalian cells as well as of recently available bioreactor technology. Growing cells according to the present invention in a bioreactor allows for large scale production and secretion of complex, fully biologically-active polypeptides into the growth media. In particular embodiments, by designing a defined media with low contents of complex proteins and using a scheme of timed-stimulation of the secretion into the media for increased titer, the purification strategy can be greatly simplified, thus lowering production cost.

1. Anchorage-dependent and non-anchorage-dependent cultures.

Animal and human cells can be propagated *in vitro* in two modes: as non-anchorage dependent cells growing freely in suspension throughout the bulk of the culture; or as anchorage-dependent cells requiring attachment to a solid substrate for their propagation (*i.e.*, a monolayer type of cell growth).

Non-anchorage dependent or suspension cultures from continuous established cell lines are the most widely used means of large scale production of cells and cell products. Large scale suspension culture based on microbial (bacterial and yeast) fermentation technology has clear advantages for the manufacturing of mammalian cell products. The processes are relatively straightforward to operate and scale up. Homogeneous conditions can be provided in the reactor which allows for precise monitoring and control of temperature, dissolved oxygen, and pH, and ensure that representative samples of the culture can be taken.

However, suspension cultured cells cannot always be used in the production of biologicals. Suspension cultures are still considered to have tumorigenic potential and thus their use as substrates for production put limits on the use of the resulting products in human and veterinary applications (Petricciani, 1985; Larsson and Litwin, 1987). Viruses propagated in suspension cultures as opposed to anchorage-dependent cultures can sometimes cause rapid changes in viral markers, leading to reduced immunogenicity (Bahnemann, 1980). Finally, sometimes even recombinant cell lines can secrete considerably higher amounts of products when propagated as anchorage-dependent cultures as compared with the same cell line in suspension (Nilsson and Mosbach, 1987). For these reasons, different types of anchorage-dependent cells are used extensively in the production of different biological products.

The current invention includes cells which are anchorage-dependent of nature. Anchorage-dependent cells, when grown in suspension, will attach to each other and grow in clumps, eventually suffocating cells in the inner core of each clump as they reach a size that leaves the core cells unsustainable by the culture conditions. Therefore, an efficient means of

large-scale culture of anchorage-dependent cells is also provided in order to effectively take advantage of the cells' capacity to secrete heterologous proteins.

2. Reactors and processes for suspension.

Large scale suspension culture of mammalian cultures in stirred tanks is contemplated. The instrumentation and controls for bioreactors have been adapted, along with the design of the fermentors, from related microbial applications. However, acknowledging the increased demand for contamination control in the slower growing mammalian cultures, improved aseptic designs have been implemented, improving dependability of these reactors. Instrumentation and controls include agitation, temperature, dissolved oxygen, and pH controls. More advanced probes and autoanalyzers for on-line and off-line measurements of turbidity (a function of particles present), capacitance (a function of viable cells present), glucose/lactate, carbonate/bicarbonate and carbon dioxide are also available. Maximum cell densities obtainable in suspension cultures are relatively low at about $2-4 \times 10^6$ cells/ml of medium (which is less than 1 mg dry cell weight per ml), well below the numbers achieved in microbial fermentation.

Two suspension culture reactor designs are most widely used in the industry due to their simplicity and robustness of operation - the stirred reactor and the airlift reactor. The stirred reactor design has successfully been used on a scale of 8000 liter capacity for the production of interferon (Phillips *et al.*, 1985; Mizrahi, 1983). Cells are grown in a stainless steel tank with a height-to-diameter ratio of 1:1 to 3:1. The culture is usually mixed with one or more agitators, based on bladed disks or marine propeller patterns. Agitator systems offering less shear forces than blades have been described. Agitation may be driven either directly or indirectly by magnetically coupled drives. Indirect drives reduce the risk of microbial contamination through seals on stirrer shafts.

The airlift reactor, also initially described for microbial fermentation and later adapted for mammalian culture, relies on a gas stream to both mix and oxygenate the culture. The gas stream enters a riser section of the reactor and drives circulation. Gas disengages at the culture surface, causing denser liquid free of gas bubbles to travel downward in the downcomer section

of the reactor. The main advantage of this design is the simplicity and lack of need for mechanical mixing. Typically, the height-to-diameter ratio is 10:1. The airlift reactor scales up relatively readily, has good mass transfer of gasses and generates relatively low shear forces.

5 Most large-scale suspension cultures are operated as batch or fed-batch processes because they are the most straightforward to operate and scale up. However, continuous processes based on chemostat or perfusion principles are available.

10 A batch process is a closed system in which a typical growth profile is seen. A lag phase is followed by exponential, stationary and decline phases. In such a system, the environment is continuously changing as nutrients are depleted and metabolites accumulate. This makes analysis of factors influencing cell growth and productivity, and hence optimization of the process, a complex task. Productivity of a batch process may be increased by controlled feeding of key nutrients to prolong the growth cycle. Such a fed-batch process is still a closed system because cells, products and waste products are not removed.

20 In what is still a closed system, perfusion of fresh medium through the culture can be achieved by retaining the cells with a fine mesh spin filter and spinning to prevent clogging. Spin filter cultures can produce cell densities of approximately 5×10^7 cells/ml. A true open system and the most basic perfusion process is the chemostat in which there is an inflow of medium and an outflow of cells and products. Culture medium is fed to the reactor at a predetermined and constant rate which maintains the dilution rate of the culture at a value less than the maximum specific growth rate of the cells (to prevent washout of the cell mass from the reactor). Culture fluid containing cells, cell products and byproducts is removed at the same rate.
25 These perfused systems are not in commercial use for production from mammalian cell culture.

3. Non-perfused attachment systems.

Traditionally, anchorage-dependent cell cultures are propagated on the bottom of small glass or plastic vessels. The restricted surface-to-volume ratio offered by classical and traditional
30 techniques, suitable for the laboratory scale, has created a bottleneck in the production of cells

and cell products on a large scale. To provide systems that offer large accessible surfaces for cell growth in small culture volume, a number of techniques have been proposed: the roller bottle system, the stack plates propagator, the spiral film bottles, the hollow fiber system, the packed bed, the plate exchanger system, and the membrane tubing reel. Since these systems are non-homogeneous in their nature, and are sometimes based on multiple processes, they can sometimes have limited potential for scale-up, difficulties in taking cell samples, limited potential for measuring and controlling the system and difficulty in maintaining homogeneous environmental conditions throughout the culture.

A commonly used process of these systems is the roller bottle. Being little more than a large, differently shaped T-flask, simplicity of the system makes it very dependable and, hence, attractive. Fully automated robots are available that can handle thousands of roller bottles per day, thus eliminating the risk of contamination and inconsistency associated with the otherwise required intense human handling. With frequent media changes, roller bottle cultures can achieve cell densities of close to 0.5×10^6 cells/cm² (corresponding to 10^9 cells/bottle or 10^7 cells/ml of culture media).

4. Cultures on microcarriers

Van Wezel (1967) developed the concept of the microcarrier culturing systems. In this system, cells are propagated on the surface of small solid particles suspended in the growth medium by slow agitation. Cells attach to the microcarriers and grow gradually to confluency of the microcarrier surface. In fact, this large scale culture system upgrades the attachment dependent culture from a single disc process to a unit process in which both monolayer and suspension culture have been brought together. Thus, combining the necessary surface for the cells to grow with the advantages of the homogeneous suspension culture increases production.

The advantages of microcarrier cultures over most other anchorage-dependent, large-scale cultivation methods are several fold. First, microcarrier cultures offer a high surface-to-volume ratio (variable by changing the carrier concentration) which leads to high cell density yields and a potential for obtaining highly concentrated cell products. Cell yields are up to $1-2 \times 10^7$ cells/ml

when cultures are propagated in a perfused reactor mode. Second, cells can be propagated in one unit process vessels instead of using many small low-productivity vessels (*i.e.*, flasks or dishes). This results in far better utilization and a considerable saving of culture medium. Moreover, propagation in a single reactor leads to reduction in need for facility space and in the number of handling steps required per cell, thus reducing labor cost and risk of contamination.

Third, the well-mixed and homogeneous microcarrier suspension culture makes it possible to monitor and control environmental conditions (*e.g.*, pH, pO₂, and concentration of medium components), thus leading to more reproducible cell propagation and product recovery.

Fourth, it is possible to take a representative sample for microscopic observation, chemical testing, or enumeration. Fifth, since microcarriers settle out of suspension easily, use of a fed-batch process or harvesting of cells can be done relatively easily. Sixth, the mode of the anchorage-dependent culture propagation on the microcarriers makes it possible to use this system for other cellular manipulations, such as cell transfer without the use of proteolytic enzymes, cocultivation of cells, transplantation into animals, and perfusion of the culture using decanters, columns, fluidized beds, or hollow fibers for microcarrier retainment. Seventh, microcarrier cultures are relatively easily scaled up using conventional equipment used for cultivation of microbial and animal cells in suspension.

5. Microencapsulation of mammalian cells

One method which has shown to be particularly useful for culturing mammalian cells is microencapsulation. The mammalian cells are retained inside a semipermeable hydrogel membrane. A porous membrane is formed around the cells permitting the exchange of nutrients, gases, and metabolic products with the bulk medium surrounding the capsule. Several methods have been developed that are gentle, rapid and non-toxic and where the resulting membrane is sufficiently porous and strong to sustain the growing cell mass throughout the term of the culture. These methods are all based on soluble alginate gelled by droplet contact with a calcium-containing solution. Lim (U.S. Patent 4,321,883) describes cells concentrated in an approximately 1% solution of sodium alginate which are forced through a small orifice, forming droplets, and breaking free into an approximately 1% calcium chloride solution. The droplets are

then cast in a layer of polyamino acid that ionically bonds to the surface alginate. Finally the alginate is reliquified by treating the droplet in a chelating agent to remove the calcium ions. Other methods use cells in a calcium solution to be dropped into a alginate solution, thus creating a hollow alginate sphere. A similar approach involves cells in a chitosan solution dropped into alginate, also creating hollow spheres.

Microencapsulated cells are easily propagated in stirred tank reactors and, with beads sizes in the range of 150-1500 μm in diameter, are easily retained in a perfused reactor using a fine-meshed screen. The ratio of capsule volume to total media volume can kept from as dense as 1:2 to 1:10. With intracapsular cell densities of up to 10^8 , the effective cell density in the culture is $1-5 \times 10^7$.

The advantages of microencapsulation over other processes include the protection from the deleterious effects of shear stresses which occur from sparging and agitation, the ability to easily retain beads for the purpose of using perfused systems, scale up is relatively straightforward and the ability to use the beads for implantation.

6. Perfused attachment systems

Perfusion refers to continuous flow at a steady rate, through or over a population of cells (of a physiological nutrient solution). It implies the retention of the cells within the culture unit as opposed to continuous-flow culture which washes the cells out with the withdrawn media (*e.g.*, chemostat). The idea of perfusion has been known since the beginning of the century, and has been applied to keep small pieces of tissue viable for extended microscopic observation. The technique was initiated to mimic the cells milieu *in vivo* where cells are continuously supplied with blood, lymph, or other body fluids. Without perfusion, cells in culture go through alternating phases of being fed and starved, thus limiting full expression of their growth and metabolic potential. The current use of perfused culture is to grow cells at high densities (*i.e.*, $0.1-5 \times 10^8$ cells/ml). In order to increase densities beyond $2-4 \times 10^6$ cells/ml (or 2×10^5 cells/cm²), the medium has to be constantly replaced with a fresh supply in order to make up for nutritional deficiencies and to remove toxic products. Perfusion allows for a far better control of

the culture environment (pH, pO₂, nutrient levels, *etc.*) and is a means of significantly increasing the utilization of the surface area within a culture for cell attachment.

Microcarrier and microencapsulated cultures are readily adapted to perfused reactors but, as noted above, these culture methods lack the capacity to meet the demand for cell densities above 10⁸ cells/ml. Such densities will provide for the advantage of high product titer in the medium (facilitating downstream processing), a smaller culture system (lowering facility needs), and a better medium utilization (yielding savings in serum and other expensive additives). Supporting cells at high density requires efficient perfusion techniques to prevent the development of non-homogeneity.

The cells of the present invention may, irrespective of the culture method chosen, be used in protein production and as cells for *in vitro* cellular assays and screens as part of drug development protocols.

J. Kits

All the essential materials and reagents required for the various aspects of the present invention may be assembled together in a kit. When the components of the kit are provided in one or more liquid solutions, the liquid solution preferably is an aqueous solution, with a sterile aqueous solution being particularly preferred.

For *in vivo* use, the instant compositions may be formulated into a single or separate pharmaceutically acceptable syringeable composition. In this case, the container means may itself be an inhalant, syringe, pipette, eye dropper, or other such like apparatus, from which the formulation may be applied to an infected area of the body, such as the lungs, injected into an animal, or even applied to and mixed with the other components of the kit.

The components of the kit may also be provided in dried or lyophilized forms. When reagents or components are provided as a dried form, reconstitution generally is by the addition of a suitable solvent. It is envisioned that the solvent also may be provided in another container

means. The kits of the invention may also include an instruction sheet defining administration of the gene therapy and/or the chemotherapeutic drug.

The kits of the present invention also will typically include a means for containing the vials in close confinement for commercial sale such as, *e.g.*, injection or blow-molded plastic containers into which the desired vials are retained. Irrespective of the number or type of containers, the kits of the invention also may comprise, or be packaged with, an instrument for assisting with the injection/administration or placement of the ultimate complex composition within the body of an animal. Such an instrument may be an inhalant, syringe, pipette, forceps, measured spoon, eye dropper or any such medically approved delivery vehicle. Additionally, instructions for use of the kit components is typically included.

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1

Modification of the RB Protein

A. Construction of RB cDNAs Expressing N-terminal Truncated pRB Proteins

For construction of modified RB cDNAs with various N-terminal deletions, a series of PCR™ primers were designed and synthesized according to the sequences of RB cDNA. The sense primers were determined by the RB cDNA sequences downstream of the deleted N-terminal sequence. All primers contain a *Hind*III restriction site (underlined) at the 5'-end and the consensus Kozak cassette (GCCGCC) followed by an ATG (*italics*). The complete nucleotide sequences of the sense primers are as follows:

5'-CCCAAGCTTGCCGCCATGGAGCAGGACAGCGGCCCGGAC-3' (OMRbSd2-34;
SEQ ID NO:14);

5'-CCCAAGCTTGCCGCCATGGATTTTACTGCATTATGTCAG-3' (OMRbSd2-55;
SEQ ID NO:15);

5 5'-CCCAAGCTTGCCGCCATGGAGAAAGTTTCATCTTGTGAT-3' (OMRbSd2-78;
SEQ ID NO:16);

5'-CCCAAGCTTGCCGCCATGCTGTGGGGAATCTGTATCTTT-3' (OMRbSd2-97;
SEQ ID NO:17);

10 5'-CCCAAGCTTGCCGCCATGTCAAGACTGTTGAAGAAG-3' (OMRbSd1-147, SEQ
ID NO:18).

The anti-sense primer 5'-GTCCAAGAGAATTCATAAAAGG-3' (OMRbAS300; SEQ
ID NO:13) overlaps with the *EcoRI* site (underlined) at the nucleotide +900 of the RB cDNA
(the A of the first in-frame ATG is designated as position +1). The anti-sense primer was paired
with each sense primer described above to amplify various modified 5'-RB cDNA fragments
using plasmid F7 as template (which contains the full-length RB cDNA).

After amplification by PCRTM with each pair of primers, the DNA fragments were
digested with *HindIII* and *EcoRI* and subcloned into plasmid pCMVRB¹¹⁰ which had been cut
with the same enzymes. The resultant expression plasmids carrying the modified RB cDNAs
with N-terminal deletions corresponding to amino acids 2-34 (SEQ ID NO:28 (nucleic acid
sequence) and SEQ ID NO:29 (amino acid sequence)), 2-55 (SEQ ID NO:30 (nucleic acid
sequence) and SEQ ID NO:31 (amino acid sequence)), 2-78 (SEQ ID NO:32 (nucleic acid
sequence) and SEQ ID NO:33 (amino acid sequence)), 2-97 (SEQ ID NO:34 (nucleic acid
sequence) and SEQ ID NO:35 (amino acid sequence)) and 1-147 (SEQ ID NO:36 (nucleic acid
sequence) and SEQ ID NO:37 (amino acid sequence)) were named as pCMVRBd₂₋₃₄ (a deletion
of amino acids 2 to 34 of the wild type RB protein), pCMVRBd₂₋₅₅ (a deletion of amino acids 2
to 55 of the wild type RB protein), pCMVRBd₂₋₇₈ (a deletion of amino acids 2 to 78 of the wild
type RB protein), pCMVRBd₂₋₉₇ (a deletion of amino acids 2 to 97 of the wild type RB protein)
and pCMVRBd₁₋₁₄₇ (a deletion of amino acids 1 to 147 of the wild type RB protein; amino acid
148 is a methionine) respectively.

B. Construction of RB cDNAs with Internal Deletions or Mutations

A total of seven pRB expression plasmids carrying RB cDNAs with varying internal deletions or mutations have been constructed, namely pCMVRBd₃₁₋₁₀₇ (a deletion of amino acids 31 to 107 of the wild type RB protein), pCMVRBd₇₇₋₁₀₇ (a deletion of amino acids 77 to 107 of the wild type RB protein), pCMVRBm_{111/112} (a mutation of amino acid 111 of the wild type RB protein from aspartic acid to glycine and a mutation of amino acid 112 from glutamic acid to aspartic acid), pCMVRBd₁₁₁₋₁₈₁ (a deletion of amino acids 111 to 181 of the wild type RB protein), pCMVRBd₁₁₁₋₂₄₁ (a deletion of amino acids 111 to 241 of the wild type RB protein), pCMVRBd₁₈₁₋₂₄₁ (a deletion of amino acids 181 to 241 of the wild type RB protein) and pCMVRBd₂₄₂₋₃₀₀ (a deletion of amino acids 242 to 300 of the wild type RB protein).

For the construction of pCMVRBd₃₁₋₁₀₇, an RB cDNA fragment from nucleotide position +325 to +910 was amplified from the plasmid F7 by PCRTM using the primers 5'-GCGCCTGAGGACCTAGATGAGATGTCGTTC-3' (SEQ ID NO:19) and OMRbAS300 (SEQ ID NO:13). This RB cDNA fragment was digested with *Bsu*36I (underlined) and *Eco*RI (from OMRbAS300), and inserted into plasmid pCMVRB¹¹⁰ digested with the same enzymes, to replace the original RB cDNA fragment from nucleotides +91 to +900. The nucleic acid sequence of pRBΔ31-107 is SEQ ID NO:38, and the corresponding amino acid sequence is SEQ ID NO:39.

For the construction of pCMVRBd₇₇₋₁₀₇, an RB cDNA fragment (nucleotides +328 to +910) was amplified from the plasmid F7 by PCRTM using the oligonucleotides 5'-GCGGTAAACCCTAGATGAGATGTCGTTCACT-3' (SEQ ID NO:20) and OMRbAS300 (SEQ ID NO:13), followed by digestion with *Hpa*I (underlined) and *Eco*RI. The amplified, digested fragment was inserted into plasmid pCMVRB¹¹⁰ digested with the same enzymes, to replace the RB cDNA fragment from nucleotides +230 to +900. The nucleic acid sequence of pRBΔ77-107 is SEQ ID NO:40, and the corresponding amino acid sequence is SEQ ID NO:41.

For the construction of pCMVRBm_{111/112}, two pairs of primers were used to change nucleotide A (position +332 of the wild-type RB cDNA) to G, in order to change the codon for aspartic acid (GAT) to glycine (GGT), thus creating a new restriction enzyme site, *Avr*II, and nucleotide G (position +336 of the wild-type RB cDNA) to T, in order to change the codon for glutamic acid (GAG) to aspartic acid (GAT). The first pair of primers are 5'-CCCAAGCTTGCCGTCATGCCGCCCAAACCCCCGA-3' (OMRBS1; SEQ ID NO:21) and 5'-CTCACCTAGGTCAACTGCTGCAAT-3' (OMRbAS332; SEQ ID NO:22; the mutated base is in bold). The second pair of primers are 5'-GTTGACCTAGGTGATATGTCGTTC-3' (OMRbS332; SEQ ID NO:23; the mutated bases are in bold) and OMRbAS300 (SEQ ID NO:13). The PCRTM products amplified with OMRBS1 and OMRbAS332 were digested with *Hind* III and *Avr*II (underlined), and those amplified with OMRbS332 and OMRbAS300 were digested with *Avr*II and *Eco*RI. These fragments were ligated together into plasmid pCMVRB¹¹⁰ digested with *Hind*III and *Eco*RI to replace the corresponding wild-type RB cDNA sequences. The nucleic acid sequence of pRBm111/112 is SEQ ID NO:50, and the corresponding amino acid sequence is SEQ ID NO:51.

For the construction of pCMVRBd₁₁₁₋₁₈₁, the RB cDNA fragment (nucleotides +543 to +910) was amplified from plasmid F7 by PCRTM using the oligonucleotides 5'-GCGCCTAGGATCTACTGAAATAAATTCTGCA-3' (SEQ ID NO:24) and OMRbAS300 (SEQ ID NO:13), followed by digestion with *Avr*II (underlined) and *Eco*RI. This fragment was then ligated into pCMVRBm_{111/112} (above) digested with the same enzymes to replace the RB cDNA fragment from nucleotides +331 to +900. The nucleic acid sequence of pRBΔ111-181 is SEQ ID NO:42, and the corresponding amino acid sequence is SEQ ID NO:43.

For the construction of pCMVRBd₁₁₁₋₂₄₁, a 5' RB cDNA fragment containing nucleotides +1 to +331 was obtained by digestion of pCMVRBm₁₁₁ with *Hind*III and *Avr*II. The 3' RB cDNA fragment beginning from nucleotide +722 was isolated from the same plasmid digested with *Pvu*II and *Bam*HI. Then the two DNA fragments (in-frame) were ligated into pCMV-G digested with *Hind*III and *Bam*HI. The nucleic acid sequence of pRBΔ111-241 is SEQ ID NO:44, and the corresponding amino acid sequence is SEQ ID NO:45.

For the construction of pCMVRBd₁₈₁₋₂₄₁, a 5'-RB cDNA fragment containing nucleotides +1 to +538 was amplified from plasmid F7 by PCRTM with primers OMRBS1 (SEQ ID NO:21) and 5'-CCCGATATCAACTGCTGGGTTGTGTCAAATA-3' (SEQ ID NO:25) using plasmid F7 as a template. The obtained RB cDNA fragment was cut with *Hind*III and *Eco*RV (underlined), and inserted into pCMVRB¹¹⁰ to replace the original 5' RB cDNA fragment between the *Hind*III and *Pvu*II sites. The nucleic acid sequence of pRBΔ181-241 is SEQ ID NO:46, and the corresponding amino acid sequence is SEQ ID NO:47.

For the construction of pCMVRBd₂₄₂₋₃₀₀, primers OMRBS1 (SEQ ID NO:21) and 5'-CCCGAATTCGTTTTATATGGTTCTTTGAGCAA-3' (SEQ ID NO:26) were used to amplify the 5' RB cDNA fragment containing nucleotides +1 to +722 using plasmid F7 as a template. The amplified product was digested with *Hind*III and *Eco*RI (underlined), and inserted into pCMVRB¹¹⁰ digested with the same enzymes to replace the original 5' RB cDNA sequences from nucleotides +1 to +900. The nucleic acid sequence of pRBΔ242-300 is SEQ ID NO:48, and the corresponding amino acid sequence is SEQ ID NO:49.

C. Characterization of N-terminal Modified RB Proteins

An RB-defective bladder carcinoma cell line, 5637 was transfected with the expression plasmids carrying the modified RB cDNAs driven by a CMV promoter. The biological function of the mutant pRBs was evaluated by a combined technique involving immunocytochemical staining and [³H]-thymidine *in situ* labeling of the tumor cells after transfection (Xu *et al.*, 1994a; 1994b).

Tumor cells were seeded onto coverslips in medium containing tetracycline and transfected with plasmids expressing pRB⁹⁴, pRB¹¹⁰ or other mutant RB proteins. At specified time point after removal of tetracycline from the culture medium, the cells were incubated with 1 ml of fresh medium containing 10 μCi [³H]-methyl thymidine (Amersham, Arlington Heights, IL) for 2 hours at 37°C, then fixed and immunochemically stained for expression of RB protein as described previously (Xu *et al.*, 1991a; 1991b). Stained slides were subsequently coated with

a thin layer of gelatin and dried at 37°C overnight. The slides were then overlaid with autoradiographic emulsion (Type NTB2, Eastman Kodak, Rochester, NY) and exposed for 2 days. After development, slides were examined under a light microscope. Twenty-four hours after transfection, cells were processed for immunocytochemical staining of RB protein and [³H]-thymidine incorporation assay as described above.

The results are illustrated in Table 5. When up to 55 amino acid residues were deleted from the N-terminal of pRB, the DNA synthesis was not significantly reduced in the cells transfected with the mutant pRB expression plasmids compared to cells expressing the full-length RB protein. However, when another 23 amino acids were removed from the N-terminal, the cellular DNA synthesis was dramatically suppressed by expression of the truncated pRB.

Table 5
% Cells Incorporating [³H]-Thymidine

<u>RB Construct</u>	<u>RB⁺</u>	<u>RB⁻</u>
Wild-Type	14	41
d2-34	12	42
d2-55	11	43
d2-78	3	41
d2-97	3	42
d1-112 (RB ⁹⁴)	2	42
d1-147	4	42
d31-107	3	41
d77-107	2	40
d111-112	6	40
d111-181	3	38
d111-241	2	40
d111-414	24	42
d181-241	8	43
d242-300	17	43

As demonstrated in Table 5, the pRB mutants with any deletions between amino acid 55 and 181 significantly inhibit DNA synthesis after being introduced into the tumor cells. Of note, cells transfected with pRBs containing deletions only between amino acid 181 and 241 showed weaker inhibition of DNA synthesis than those transfected with plasmids expressing pRBs carrying deletions between amino acid 55 and 181, although these were still more effective than cells transfected with the full-length pRB expression plasmid. Thus, in view of this data, modifications that combine certain of the above deletions, for example a deletion between amino acid 1 and amino acid 241, would be expected to have similar significant DNA synthesis inhibitory activity.

Additionally, two pRB mutants with two deletions each, either between amino acid 2 and 34 and between amino acids 76 and 112, or between amino acids 2 and 55 and between amino acids 76 and 112 significantly inhibited DNA synthesis as compared to the wild-type RB. The results indicated the boundary of the putative N-terminal domain probably located between amino acid 182 and 300, most probably between amino acid 182 and 241. In addition, a pRB carrying a point mutation at amino acid position 111 converting aspartic acid to glycine significantly suppressed DNA synthesis, further suggesting that this region is vital for regulating pRB function.

EXAMPLE 2

Modification of the CMV Promoter/Enhancer Controlling Expression of the VP16

Transactivating Domain in the Tetracycline-Responsive Gene Expression System

The modified retinoblastoma genes and proteins described above have a number of practical utilities, including, but not limited to, gene therapy. For these aspects, expression systems are needed. While systems such as those described above are appropriate for certain embodiments, they have certain shortcomings in relation to gene therapy using cytotoxic constructs. The original tetracycline-responsive gene expression system of Gossen and Bujard (1992) is an attractive system, but has certain drawbacks, such as squelching effects on cell

growth (Gill and Ptashne, 1988). To overcome these and other drawbacks, the inventors have improved the tetracycline-responsive gene expression system.

The original tetracycline repressor/operator-based regulatory system consists of two plasmids, pUHD15-1 and pUHC13-3 (U. S. Patent 5,464,758, incorporated in its entirety herein by reference; Gossen and Bujard 1992). pUHC13-3 is a tetracycline (Tc; tet) sensitive expression vector containing a hybrid minimal human CMV promoter, in which tet operator sequences had been inserted upstream of the TATA box. pUHD15-1 contains sequences encoding a tetracycline responsive transactivator (tTA), with expression driven by a wild-type CMV promoter. In transient experiments using this system, the inventors found that efficiently reversible transgene expression was observed in many tumor cell lines studied. However, attempts to isolate long-term clones expressing the reporter gene in a tetracycline-responsive manner were unsuccessful. This was most likely caused by the high intracellular levels of the tTA transactivator, whose expression was driven by the strong CMV promoter/enhancer sequence in the plasmid pUHD15-1. The tTA transactivator contains the VP-16 activating domain, which is known to have squelching effects on cell growth (Gill and Ptashne, 1988).

Therefore, to resolve this problem and to further improve the system, the tTA expression cassette was first modified by replacing the strong CMVp enhancer (Boshart *et al.*, 1985) in the original pUHD15-1 plasmid with a pair of 19 bp imperfect direct repeat sequence (a portion of the CMVp enhancer; SEQ ID NO:5). The modification of the hCMV promoter/enhancer was done by removal of a portion of the 5' enhancer sequences from the hCMV promoter.

Three pairs of oligonucleotide primers were designed based on the published sequence of the hCMV promoter (Boshart *et al.*, 1985). A *Xho*I and an *Eco*RI restriction enzyme site (underlined) was added to the 5' end of each sense and the anti-sense oligo, respectively. The sense oligos are: 5'-CCGCTCGAGCAATGGGCGTGATAGCGG-3' (OMCMVs1; SEQ ID NO:6); 5'-CCGCTCGAGCACCAAAATCAACGGGA-3' (OMCMVs2; SEQ ID NO:7) and 5'-CCGCTCGAGCAACTCCGCCCCATTGAC-3' (OMCMVs3; SEQ ID NO:8), respectively, and

they shared the same anti-sense primer, 5'-TAGACATATGAATTCGCGGCC-3' (OMCMVas; SEQ ID NO:9).

The template used in PCRTM amplification was plasmid pUHD15-1. PCRTM amplification with primer pairs of OMCMVs1 + OMCMVas; OMCMVs2 + OMCMVas and OMCMVs3 + OMCMVas, generated three shorter versions of CMV promoter with lengths of 282 bp (namely mhCMVp1), 203 bp (mhCMVp2) and 168 bp (mhCMVp3) respectively. The purified shortened CMV promoter/enhancer fragments were double digested with *Xho*I and *Eco*RI, and inserted into pUHD15-1 to replace the original hCMV promoter. This produced three new tTA expressing plasmids, namely pmCMV1-tTA, pmCMV2-tTA and pmCMV3-tTA.

To determine the relative strength of these promoters, the tTA in these newly constructed plasmids, as well as plasmid pUHD15-1, was replaced by a chloramphenicol acetyltransferase (CAT) gene from plasmid pRc/CMV-CAT (Invitrogen, San Diego, CA), thus generating four CAT expression plasmids, pmCMV1-CAT, pmCMV2-CAT, pmCMV3-CAT and pCMV-CAT. In these plasmids, CAT expression is driven by mhCMVp1, mhCMVp2, mhCMVp3 and the full-length hCMVp, respectively. To evaluate the relative activity of the modified CMV promoters, the CAT expression plasmids were introduced into three cell lines, the tumor cell lines 5637 and Saos2, and the embryonal kidney cell line 293, *via* the Lipofectin method (Life Technologies, Gaithersburg, MD). Forty-eight hours after transfection, cell lysates were prepared and CAT activity was measured by a CAT FLASH assay kit from Stratagene (Stratagene, La Jolla, CA).

As shown in FIG. 1, after enhancer sequences were partially removed, the activity of the promoter was dramatically reduced in all three transfected cell lines. FIG. 1 is a graphical representation of the CAT activity in the 5637 and Saos-2 cell lines. The more enhancer sequences that were deleted, the weaker was the promoter that remained. The order of promoter activity from strongest to weakest is hCMV, mhCMVp1, mhCMVp2 and mhCMVp3. The activity of mhCMVp1 is 17.7% of the full-length hCMV promoter, while the mhCMVp3 activity is only 3.3% of the hCMV promoter in 5637 cells (FIG. 1). After comparing the relative promoter activity of the modified promoters, mhCMVp1 (SEQ ID NO:5) was chosen for the

modified tetracycline regulatable gene expression system. mhCMVp1 showed optimal tetracycline-controlled transactivator (tTA) expression with no squelching effects on host cell growth (FIG. 2), an important characteristic for potential use in human gene therapy.

EXAMPLE 3

Construction of Single Plasmid, Tetracycline-Regulated Vector

A single plasmid vector named EC1214A was constructed. This plasmid contains: 1) the modified tetracycline-responsive transactivator (tTA) expression cassette to eliminate the squelching effects of tTA on host cell growth; 2) the tTA-dependent promoter from plasmid pUHC13-3; 3) a generic intron sequence; 4) a multiple cloning site downstream of the promoter and intron; and 5) a neo^R expression cassette to allow G418 selection. Expression in this system is regulated by tetracycline, or a tetracycline analog. A "tetracycline analog" will be understood to be any one of a number of compounds that are closely related to tetracycline, and which bind to the tet repressor with at least an affinity (K_a) of at least $10^6/M$, preferably with a K_a of $10^9/M$, and more preferably with a K_a of $10^{11}/M$. Exemplary, but in no way limiting, of such tetracycline analogs are those disclosed by Hlavka and Boothe (1985), Mitschef (1978), the Noyee Development Corporation (1969), Evans (1968) and Dowling (1955), each of which is incorporated herein in its entirety.

Plasmid pMLSIS.CAT (Choi *et al.*, 1991) contains an generic intron sequence which consists of a portion of the 5'-untranslated leader from the adenovirus-major-late region, which contains part of the first exon of the tripartite and the first intervening sequence, as well as a synthetic splice donor/acceptor sequence derived from an IgG variable region. A pair of oligonucleotides, 5'-CTAGAAATTCGCTGTCTGCG-3' (SEQ ID NO:10) and 5'-GCTCTAGATGCAGTTGGACCTGGGAG-3' (SEQ ID NO:11), flanking the intron sequence in plasmid pMLSIS.CAT and containing an *Eco*RI and *Xba*I site, respectively (underlined), were synthesized. After amplification by PCR[™], the intron fragment was digested with *Eco*RI and *Xba*I, and inserted into the corresponding enzyme sites in plasmid pUHD15-1.

Subsequently, a small DNA fragment containing *Clal*, *HindIII*, *EcoRV*, *EcoRI*, *PstI*, *SmaI* and *BamHI* cloning sites (obtained from plasmid pBluescriptSK) was inserted into the new plasmid downstream of the intron to produce an expression vector containing the hCMV promoter, a generic intron, multiple cloning sites and a polyadenylation signal from the SV40 virus. This intermediate vector was given the name of pCMV-G. The SV40 polyadenylation signal of pCMV-G was then replaced by a HSV thymidine kinase (TK) gene polyadenylation signal sequence to generate a plasmid, named pCMV*-G-TKpA.

Plasmid pRc/CMV (Invitrogen, San Diego, CA) was double digested with restriction enzymes *NruI* and *XbaI*. The 5' overhang from the *XbaI* digest was filled in by Klenow fragment of DNA polymerase (Life Technologies, Gaithersburg, MD), and the blunt-ended insert was ligated to a DNA fragment containing mhCMV1-tTA obtained from plasmid pmCMV1-tTA (Example 2). The new plasmid was named pmCMV1-tTA.neo.

Finally, a DNA fragment containing the tTA-dependent promoter, the generic intron and the TK polyadenylation signal was isolated from plasmid pCMV*-G-TKpA, and inserted into the *Bgl/II* site of plasmid pmCMV1-tTA.neo to produce a vector named EC1214A, which carries both the tTA expression cassette and the tTA-dependent promoter as well as a selection marker, the neomycin resistance gene.

EXAMPLE 4

Construction of a Single Plasmid Tetracycline Positively-Induced (Tet-on) Vector

The original tetracycline repressor/operator-based tet-on system also consists of two plasmids, pUHD17-1neo (or pUHD172-1neo) and pUHC13-3 (Gossen *et al.*, 1995). pUHC13-3 is a tetracycline sensitive expression vector containing a hybrid minimal human CMV promoter, in which tet operator sequences had been inserted upstream of the TATA box. pUHD17-1neo or pUHD172-1neo contains sequences encoding a reverse tetracycline responsive transactivator (rtTA), with expression driven by a wild-type CMV promoter. In transient experiments using this system, it was found that efficiently reversible transgene expression was observed in many tumor cell lines studied. As opposed to the original tetracycline system, expression is turned on

in the presence of tetracycline or a tetracycline analog, such as doxycycline, while expression is turned off in the absence of tetracycline. However, the rtTA transactivator contains the VP-16 activating domain, which is known to have squelching effects on cell growth (Gill and Ptashne, 1988).

5

10

Therefore, to resolve this problem and to further improve the system, the rtTA expression cassette was first modified by replacing the strong CMVp enhancer (Boshart *et al.*, 1985) in the pUHD17-1neo or pUHD172-1neo plasmid with a pair of 19 bp imperfect direct repeat sequence (SEQ ID NO:5). The modification of the hCMV promoter/enhancer was done by removal of a portion of the 5' enhancer sequences from the hCMV promoter (Example 2). The new rtTA expressing plasmid was named pmCMV1-rtTA.

15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995

A single plasmid vector named EC1214B was constructed using pmCMV1-rtTA. This plasmid contains: 1) the modified reverse tetracycline-responsive transactivator (rtTA) expression cassette to eliminate the squelching effects of rtTA on host cell growth; 2) the rtTA-dependent promoter from plasmid pUHC13-3; 3) a generic intron sequence; 4) a multiple cloning site downstream of the promoter and intron; and 5) a neo^R expression cassette to allow G418 selection. The construction was performed as outlined in Example 3.

EXAMPLE 5

Construction of Retinoblastoma (RB) and p53 Tetracycline-Controlled Vectors

A. Construction of Inducible pRB¹¹⁰ Expression Vector

25

To construct an inducible pRB¹¹⁰ expression plasmid, plasmid F7 (Takahashi *et al.*, 1991) or p4.95BT (Friend *et al.*, 1987), containing the full-length RB¹¹⁰ gene cDNA, was digested with the restriction enzymes *AcyI* at nucleotide -322 and *ScaI* at +3230 (the A of the second in-frame ATG start codon was designated nucleotide +19). The 5' overhangs generated by the *AcyI* digest were treated with *E. coli* DNA polymerase I in the presence of all four dNTPs to generate blunt ends. *BamHI* linkers were ligated onto the fragment, and the fragment was then digested with *BamHI* to remove excess linkers and generate *BamHI* ends (Maniatis *et al.*, 1989; Ausubel *et al.*,

1992). The resultant RB cDNA fragment of 3552 bp was inserted into the unique *Bam*HI site of EC1214A to generate pCMV*-tTA-RB¹¹⁰.

B. Construction of Inducible pRB⁹⁴ Expression Vector

It is known that the primary sequence surrounding the AUG codon GCC(^A_G)CCAAUGG (SEQ ID NO:27) is the optimal context for initiation of translation in higher eukaryotes (Kozak 1991). A surprising realization is that, although nearly all vertebrate mRNAs have features that ensure the fidelity of initiation, many mRNAs that encode critical regulatory proteins do not appear to be designed for efficient translation (Kozak 1991). In reviewing the RB cDNA sequence, it was found that the AUG start codon for both the full length pRB¹¹⁰ and the N-terminal truncated pRB⁹⁴ are in a suboptimal context for initiation of translation in higher eukaryotes. For example, there is an out-of-frame AUG codon at the nucleotide -5 position (the A of the ATG start codon for the pRB⁹⁴ cDNA is designated nucleotide +1), and the leading sequence of the ATG codon for pRB⁹⁴ is suboptimal as compared to the consensus initiator context shown above. To improve the translation efficiency of the pRB⁹⁴ cDNA, site-directed mutagenesis was used to optimize the DNA sequence upstream of the second internal in-frame ATG codon of RB⁹⁴ for optimal translational initiation.

The modified 5'-RB⁹⁴ cDNA fragment was obtained by PCRTM using plasmid F7 carrying the full-length RB¹¹⁰ cDNA as the template. The sense primer used for the PCRTM reaction (5'-CCCAAGCTTGCCGCCATGTCGTTCACTTTTAC-3'; SEQ ID NO:12) contained a *Hind*III restriction site (underlined) and a Kozak cassette (italics; Kozak, 1987). The antisense primer 5'-GTCCAAGAGAAATTCATAAAAGG-3' (OMRbAS300; SEQ ID NO:13) overlapped with the *Eco*RI site (underlined) at nucleotide +900 of the RB cDNA (the A of the first in-frame ATG is designated as position +1). The PCRTM product was digested with *Hind*III and *Eco*RI, then ligated with a DNA fragment containing the 3'-RB cDNA fragment between *Eco*RI (position +900) and *Bam*HI (+3548) isolated from plasmid F7. The entire RB⁹⁴ cDNA fragment was inserted into the *Hind*III and *Bam*HI sites of EC1214A to produce the inducible pRB⁹⁴ expression plasmid, pCMV*-tTA-RB⁹⁴.

C. Construction of Inducible p53 Expression Vector

A plasmid, pC53-SN3 (Baker *et al.*, 1990), containing the full length p53 gene cDNA was digested with *Bam*HI, and the fragment containing the full length p53 gene was inserted into the unique *Bam*HI site of EC1214A to generate pCMV*-tTA-p53.

EXAMPLE 6

Preparation of Long-Term Tumor Cell Clones with Tetracycline-Regulated pRB110, pRB94 or p53 Expression

The modified, single-plasmid tetracycline-responsive mammalian gene expression system has been used to obtain various stable tumor cell lines in which expression of the wild-type or the N-terminal truncated retinoblastoma (RB) tumor suppressor gene, or the p53 tumor suppressor gene can be reversibly turned on and off without detectable leakage.

A. Cell Culture

A breast carcinoma cell line, MDA-468 (HTB132) was obtained from ATCC and cultured in Leibovitz's L-15 (Life Technologies, Gaithersburg, MD) with 10% FBS (Life Technologies, Gaithersburg, MD). An osteosarcoma cell line, Saos2 was cultured in medium McCoy's 5A (Life Technologies, Gaithersburg, MD) with 15% FBS (Zhou *et al.*, 1994b). A bladder carcinoma cell line, 5637 (HTB9) obtained from ATCC was cultured with RPMI 1640 medium (Life Technologies, Gaithersburg, MD) containing 10% FBS. All cell culture media were supplemented with 0.5% penicillin/streptomycin. Saos2 and 5367 cells were incubated at 37°C in a 5% CO₂ incubator, while MDA-468 cells were cultured at 37°C without CO₂.

B. Stable Transfection

Tumor cells were transfected with the pRB¹¹⁰ and pRB⁹⁴ expression plasmids, pCMV*-tTA-RB¹¹⁰ and pCMV*-tTA-RB⁹⁴ via the Lipofectin method according to the manufacturer's instruction manual (Life Technologies, Gaithersburg, MD). During transfection and the subsequent procedures except where specified, 0.5 µg/ml of tetracycline (Sigma, St. Louis, MO) was added to the transfection and culture media. Forty-eight hours after transfection, G418 (Life

Technologies, Gaithersburg, MD.) was added to the culture media at a concentration of 300 µg/l. Two to three weeks later, single colonies were isolated by cloning rings. A duplicate culture was made for each isolated colony. While the original clone was kept in media containing 0.5 µg/ml tetracycline, the duplicate clone was cultured in the absence of tetracycline. The latter was immunochemically stained with a specific anti-RB antibody, RB-WL-1 (Xu *et al.*, 1989a). The matched RB-positive clones were subsequently maintained in medium containing tetracycline and G418 and expended for further analyses.

C. Transient Transfection

Tumor cells were seeded into 60-mm culture dishes or onto sterile coverslips at concentrations that would reach about 40% confluent next day. Twenty hours later, proper amount of plasmid DNA was mixed with Lipofectin reagent in Opti-MEM medium according to the manufacture's instruction manual (Life Technologies, Gaithersburg, MD). Cells were overlaid with the DNA-Lipofectin complex and incubated in a CO₂ incubator at 37°C overnight. Next day, fresh medium was added to replace the DNA-Lipofectin. Twenty-four or forty-eight hours later, cells were fixed for immunochemical staining or lysed for preparation of cell lysates.

D. Immunocytochemical Staining of RB Protein

Immunocytochemical staining was performed as described previously (Xu *et al.*, 1989a). For detection of RB expression, cells grown on coverslips were fixed in 45% (vol/vol) acetone/10% (wt / vol) formaldehyde/0.1 M phosphate buffer for 5 min. After being washed six times with phosphate-buffered saline, cells were blocked with 1% non-fat milk/1.5% goat serum or horse serum in phosphate buffer for 4 hours at room temperature. The RB-WL-1 anti-RB antibody or Canji's monoclonal anti-RB antibody (QED, San Diego, CA) was diluted to 2 µg/ml or 0.5 µg/ml respectively in the same solution plus 0.02% Triton X-100, and was incubated with the cell overnight. After being washed, the coverslips were processed for immunostaining with the avidin biotinylated peroxidase complex (ABC) method according to the technical manual (Vector Laboratories, Burlingame, CA).

E. Immunoblotting for pRB

Cell lysate was prepared as previously described (Xu *et al.*, 1991a; 1991b). Briefly, cultured cells in 60 mm dishes were lysed with 0.6 ml of ice-cold lysis buffer containing 100 mM NaCl, 0.2% NP-40, 0.2% sodium deoxycholate, 0.1% SDS and 50 mM Tris-HCl (pH8.0) with 50 µg/ml aprotinin and 1 mM PMSF. The cell lysate was passed through 21 gauge needle several times and clarified by centrifugation.

Direct Western immunoblotting was done as described previously (Xu *et al.*, 1991a; 1991b). Sixty micrograms of total cellular protein as determined by the Bradford protein assay (BioRad, Richmond, CA) was electrophoresed in an 8% SDS/polyacrylamide gel and electroblotted to Immobilon polyvinylidene difluoride membranes (PVDF) (Millipore, Bedford, MA). After being blocked with 4% bovine serum albumin/1% normal goat serum in Tris-buffered saline, membranes were incubated overnight with RB-WL-1 antibody at a final concentration of 0.4 µg/ml for RB detection. The blots were then probed by the ProtoBlot Western blot alkaline phosphatase system (Promega, Madison, WI).

F. Growth Curve Measurement

A crystal violet staining method was used to measure the cell growth changes in the presence or absence of tetracycline (Gillies *et al.*, 1986). Briefly, cells were seeded into 24-well plates in duplicate. In one set of the plates, cells were grown in medium containing 0.5 µg/ml tetracycline, while in duplicate plates, the same cells were cultured in non-tetracycline media. At each time point, cells were fixed with 1% glutaraldehyde in PBS and stained using 0.5% of crystal violet. After cells at all desired time points were collected, the crystal violet dye was extracted from the stained cells by incubating cells with Sorenson's solution containing 0.9% trisodium citrate, 0.02 N chloric acid and 45% ethanol (vol/vol). The extracted dyes were diluted properly with the Sorenson's solution and optical absorbencies at λ_{550} were measured. Growth curves were obtained by plotting the OD₅₅₀ against the time.

G. Soft Agar Assay

For soft agar assay, appropriate number of cells were mixed with 0.3% of agarose in complete medium containing 15% FBS and overlaid onto 0.7% base agar in a 35 mm tissue culture dish. Duplicate dishes were prepared for each individual cell clones. Cells in one dish were cultured in the medium containing 0.5 µg/ml of tetracycline and the other cultured in non-tetracycline medium. The medium was replenished every 3 days, and colonies (>50 cells) were counted after 3 weeks. Results were calculated as the average of three dishes per cell clone.

H. Tumorigenicity Test in Nude Mice

The tumorigenicity test has been described previously (Takahashi *et al.*, 1991). Two groups of athymus nude mice were set up for each cell clone to be tested. One group of mice were given regular water, while the other group was given water containing 5 mg/ml of tetracycline. A total of 5×10^6 cells from each RB¹¹⁰- or RB⁹⁴-reconstituted clone were injected subcutaneously in 0.2 ml of phosphate buffered saline into the right flank of nude mice. RB-negative parental controls including Saos2, 5637 and MDA-468 cells were injected at the identical concentration into the left flank of the same mice. Tumors were scored 4 weeks after injection.

I. Time Course Study of [³H]-Thymidine Incorporation

Cells from inducible RB-reconstituted clones were grown on sterile coverslips in medium containing tetracycline. At specified time point after removal of tetracycline from the culture medium, the cells were incubated with 1 ml of fresh medium containing 10 µCi [³H]-methyl thymidine (Amersham, Arlington Heights, IL) for 2 hours at 37°C, then fixed and immunochemically stained for expression of RB protein as described previously (Xu *et al.*, 1991a; 1991b). Stained slides were subsequently coated with a thin layer of gelatin and dried at 37°C overnight. The slides were then overlaid with autoradiographic emulsion (Type NTB2, Eastman Kodak, Rochester, NY) and exposed for 2 days. After development, slides were examined under a light microscope.

J. ³H]-Thymidine Incorporation of Transiently Transfected Cell Cultures

Tumor cells were seeded onto coverslips and transfected with plasmids expressing pRB⁹⁴, pRB¹¹⁰ or other mutant RB proteins. Twenty-four hours after transfection, cells were processed for immunocytochemical staining of RB protein and [³H]-thymidine incorporation assay as described in Xu *et al.* (1991b; 1991c).

K. Characterization of Long-Term Inducible RB Expression Clones

The cell growth suppression and morphological changes after RB replacement that have been reported in the literature are inconsistent. Studies done by the inventors and others indicated that replacement of the normal *RB* gene into *RB*-defective tumor cells could suppress their tumorigenic activity in nude mice (Goodrich and Lee 1993, Bookstein *et al.*, 1990a; 1990b; Chen *et al.*, 1992; Goodrich *et al.*, 1992b; Huang *et al.*, 1988 ; Kratzke *et al.*, 1993; Madreperla *et al.*, 1991; Muncaster *et al.*, 1992; Ookawa *et al.*, 1993; Sumegi *et al.*, 1990; Takahashi *et al.*, 1991; Wang *et al.*, 1993; Xu *et al.*, 1996; Xu *et al.*, 1991c; Zhou *et al.*, 1994b; Xu, 1996; Xu, 1995; Li *et al.*, 1996; Xu *et al.*, 1994b). The tumor cell lines studied were derived from widely disparate types of human cancers such as the retinoblastoma, osteosarcoma, carcinomas of the bladder, prostate, breast and lung (Goodrich and Lee, 1993; Xu, 1996; Xu, 1995 for review). Although it has been well documented that correction of the *RB* gene defect alone in tumor cells carrying multiple genetic alterations was sufficient to revert their malignant phenotype, it was more puzzling than it appeared at first sight (Klein, 1990).

As was shown in several early studies, after transient transfection with pRB-expressing plasmids, some types of the *RB*-defective tumor cells in culture displayed striking changes, including cell enlargement, senescent-like phenotype and growth cessation (Templeton *et al.*, 1991; Qin *et al.*, 1992). Subsequently, it was found that, however, long-term stable clones of the *RB*-reconstituted tumor cells can be isolated that grew just as rapidly as the parental lines. Therefore, there has been a tendency in the literature to separate the inhibition of cell growth by *RB* replacement in *RB*-defective tumor cells from its tumor suppression function (Chen *et al.*, 1992; Goodrich *et al.*, 1992b; Takahashi *et al.*, 1991; Xu *et al.*, 1991b; Zhou *et al.*, 1994b; Li *et al.*, 1996).

Three RB-defective tumor cell lines were used to establish long-term inducible RB expression clones. They were the osteosarcoma cell line, Saos2, the bladder cancer cell line, 5637 and the breast cancer cell line, MDA-468. The rationale for choosing Saos2, 5637 and MDA-468 as recipient cells was that they are the RB-defective tumor cells most in use for RB-replacement studies. The tumor cells were transfected with the inducible RB¹¹⁰ expression plasmid, pCMV*-tTA-RB¹¹⁰ and the pRB⁹⁴ expression plasmid, pCMV*-tTA-RB⁹⁴ in the presence of tetracycline. After selection in 400 µg/ml of G418 for approximately 2 to 4 weeks, well separated single colonies were isolated and maintained in tetracycline containing media. A small portion of the isolated clones were cultured separately in the absence of tetracycline (Tc) for 24 to 48 hours and stained with an anti-RB antibody, RB-WL-1. Tight control of pRB protein expression in the stable clones of Tc-responsive RB-reconstituted 5637 bladder carcinoma and MDA-MB-468 breast carcinoma cells is seen.

The RB-reconstituted 5637 cells grown in the presence of 0.5 µg/ml of Tc in the culture medium are RB⁻ by immunocytochemical staining, while after removal of Tc, the pRB expression was turned on in the RB-reconstituted 5637 cells as shown by RB⁺ immunocytochemical staining. The MDA-MB-468 breast carcinoma tumor cells were also RB⁻ by immunocytochemical staining in the presence of 0.5 µg/ml of Tc in culture medium, whereas after removal of Tc, the pRB expression was turned on in the RB-reconstituted MDA-MB-468 breast carcinoma cells as shown by RB⁺ immunocytochemical staining. Note that tetracycline is an inhibitor, rather than an inducer, in this tetracycline-responsive expression system.

The minimal concentration of tetracycline required to shut off RB expression was also tested. It was found that as little as 0.1 µg/ml of tetracycline can inhibit RB expression to non-detectable level by immunostaining, indicating that the tetracycline-regulated expression system is very sensitive to tetracycline.

Additionally, it was surprisingly found that, unlike the non-regulatable, long-term *RB*-reconstituted tumor cell lines previously reported, all the long-term tumor cell clones examined irreversibly ceased growing after pRB expression was turned on in Tc-free medium (FIG. 3A, FIG. 3B and FIG. 3C). It is known in the literature that the half-life of pRB in normal and tumor cells is only 4 to 6 hours (Mihara *et al.*, 1989; Xu *et al.*, 1994b; Xu *et al.*, 1989a), and as was illustrated in FIG. 2, using the modified tetracycline-regulatable system, expression of tTA transactivator *per se* in the presence or absence of low concentration of Tc had no effect on cell growth.

The Saos2 and 5637 clones also failed to synthesize DNA, which were followed by noticeable morphological changes and finally, by cell death. The cellular morphology was markedly altered after pRB expression was induced in Tc-free medium, including cell enlargement, flattening, and lower nucleocytoplasmic ratio than cycling G1/S cells. In the case of the bladder carcinoma cell line, 5637, changes in morphology and growth rate after either transient or stable *RB*-replacement with a non-regulatable system have not been well documented in the literature (Goodrich *et al.*, 1992b; Takahashi *et al.*, 1991; Zhou *et al.*, 1994b).

In general, the phenotypes of the established Tc-regulatable *RB*⁺ tumor lines in Tc-free medium were quite similar to those documented previously for *RB* plasmid-transfected (or *RB* retrovirus vector-infected) tumor cell mass cultures (Huang *et al.*, 1988; Templeton *et al.*, 1991; Qin *et al.*, 1992). All tumor cell clones under permissive condition for pRB expression were unable to form colonies in soft agar (FIG. 4A, FIG. 4B and FIG. 4C), and were non-tumorigenic in nude mice.

To compare *RB* with another common tumor suppressor gene, *p53*, several long-term stable tumor cell clones with Tc-regulatable wild-type *p53* expression have been established from the osteosarcoma cell line, Saos-2. A similar approach as described above was used to establish the *p53*-reconstituted Saos-2 tumor cell clones. In brief, the parental Saos-2 tumor cells were transfected with the wild-type *p53*-expressing plasmid, pCMV*-tTA-*p53* (Example 5) and selected in geneticin-containing media. The initial G418-resistant mass cultures were subjected

to at least two rounds of subcloning in order to obtain stable wild-type *p53*-reconstituted clones. Because of complete deletion of the *p53* gene, the parental Saos-2 cells have no endogenous *p53*.

With this model system, it was found that induction of wild-type *p53* expression in *p53*-reconstituted Saos-2 clones did result in growth arrest of the $RB^-/p53^{null}$ tumor cells. When the Tc-regulated *p53*-reconstituted Saos-2 clones were grown in the absence of Tc, many tumor cells shrank and detached. Furthermore, as measured by DNA fragmentation assay, abundant low molecular weight DNAs were detected only in samples extracted from *p53*-reconstituted Saos-2 tumor cells under permissive condition for *p53* expression. These observations indicate that the wild-type *p53*-induced growth arrest of the $RB^-/p53^{null}$ Saos-2 tumor cells was the result of apoptotic cell death rather than replicative senescence.

Dimri *et al.* recently reported a biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. It was shown that several human senescent cells expressed a β -galactosidase, histochemically detectable at pH 6 (Dimri *et al.*, 1995). This marker, termed senescence-associated β -galactosidase (SA- β -gal), is expressed by senescent, but not pre-senescent fibroblasts. SA- β -gal was also absent from immortal cells, but was induced by genetic manipulations that reversed immortality (Dimri *et al.*, 1995). Of note, some cells, such as adult melanocytes, expressed the SA- β -gal (pH 6 activity) independent of senescence or age. Thus, SA- β -gal is not a universal marker of replicative senescence, which is not surprising.

Nevertheless, by utilizing the instant long-term tumor cell clones with tetracycline-regulatable *pRB* or *p53* expression, the SA- β -gal (pH 6 activity) provides a simple assay allowing the further characterization of the *RB*-mediated tumor cell growth cessation. The majority (>99.9%) of young (early passage) human WI-38 fibroblasts are SA- β -gal negative. In contrast, the senescent (at population doubling level greater than 52) WI-38 cells were strongly SA- β -gal positive. All tetracycline-responsive tumor cell clones examined so far were SA- β -gal negative in the presence of tetracycline (RB^-), and were SA- β -gal positive in tetracycline-free medium

(RB⁺). The intensity of SA-β-gal staining of tumor cells in RB⁺ status, however, was variable depending on tumor cell types.

Of note, although p53 reconstitution in Saos-2 (RB⁻, p53^{null}) tumor cells with either non-inducible (Chen *et al.*, 1990; Li *et al.*, 1996) or inducible system did suppress their neoplastic phenotype, the p53 reconstituted Saos-2 clones with the tetracycline-regulatable promoter were SA-β-gal negative in either presence or absence of tetracycline. Of great interest, when the p53-reconstituted Saos-2 cells were infected with recombinant adenovirus vectors expressing the wild-type pRB¹¹⁰ in Tc-free medium, the tumor cells with both wild-type p53 and pRB¹¹⁰ expression displayed more intense SA-β-gal positive staining as compared to tumor cells only expressing pRB¹¹⁰. The results imply that the mechanisms for tumor suppression by pRB and p53 were different from each other, but expression of pRB and p53 together had synergistic effects on RB-mediated tumor cell senescence.

In consideration of its potential therapeutic use, another important finding was the fact that the pRB-mediated replicative senescence (irreversible growth cessation) was tumor-specific. The young WI-38 fibroblasts at early passage infected with recombinant adenovirus vector, AdCMVpRB110 at multiplicity of infection (MOI) of 100 remained SA-β-gal negative, and they resumed a normal growth pattern about one week post-infection. Therefore pRB is a relatively safe reagents for anticancer gene therapy. In addition to therapy of advanced malignancies, the emerging RB gene therapy also may be beneficial in treating post-surgery residue tumors, superficial cancers, or premalignancies, as well as non-malignant, hyperproliferative disorders in certain circumstances (Chang *et al.*, 1995; Xu *et al.*, 1996).

L. The broad biological basis of the RB-mediated tumor suppression.

In addition to tumor cell-specific senescence and the well-known antiproliferative effects, pRB may also play a role in inhibition of angiogenesis and in elicitation of immunogenicity of tumor cells. The inventors have shown that serum-free conditioned media (CM) collected from the tetracycline-responsive, RB-reconstituted osteosarcoma and non-small cell lung carcinoma

cell lines switched from angiogenic to anti-angiogenic after removal of Tc from the cell cultures. This switch corresponded with the onset of pRB expression as determined by Western blotting and immunohistochemistry (Dawson *et al.*, 1996). The inventors have also reported that HLA class II induction by IFN- γ in the RB-defective non-small cell lung carcinoma cell line, H2009, requires reconstitution of the wild-type RB gene expression (Lu *et al.*, 1996). The class II proteins present peptides derived from proteolytically processed antigens to CD4⁺ T lymphocytes as part of the immune response. Therefore, pRB likely has a role in mediating tumor immunogenicity as well.

To determine if replacement of the retinoblastoma (*RB*) tumor suppressor gene could inhibit invasion of *RB*-defective tumor cells, studies were conducted using the Boyden chamber assay (Li *et al.*, 1996). The studies were done in a diverse group of stable *RB*-reconstituted human tumor cell lines, including those derived from the osteosarcoma and carcinomas of the bladder, breast and lung. The expression of the exogenous wild-type RB protein in these tumor cell lines was driven by either a constitutively active promoter or an inducible promoter. It was found that significantly more tumor cells from the parental *RB*-defective cell lines and the *RB*⁻ revertants than from the *RB*-reconstituted *RB*⁺ cell lines penetrated through the Matrigel in the Boyden chamber assay ($p < 0.001$, two-tailed t-test). Of note, the inhibition of invasiveness of various *RB*-defective tumor cells by *RB* replacement was apparently well correlated with suppression of their tumorigenicity *in vivo*. In contrast, although either functional *RB* or p53 re-expression effectively suppressed tumor formation in nude mice of the *RB*⁻/p53^{null} osteosarcoma cell line, Saos-2, replacement of the wild-type p53 gene had much less impact on their invasiveness as compared to the *RB* gene.

Normal human diploid cells senesce *in vitro* and *in vivo* after a limited number of cell divisions. This process known as cellular senescence is an underlying cause of aging and a critical barrier for development of human cancers. It has also been demonstrated that *RB*/p53-defective tumor cells reexpressing functional pRB alone *via* a modified tetracycline-regulated gene expression system were irreversibly growth-arrested at G0/G1 phase of the cell cycle.

These cells displayed multiple morphological changes consistent with cellular senescence and also expressed a senescence-associated β -galactosidase biomarker.

Further studies indicated that telomerase activity, which was presumably essential for an extended proliferative life-span of neoplastic cells, was repressed in the tumor cell lines after induction of pRB (but not p53) expression. These observations suggest that pRB plays a critical role in the intrinsic cellular senescence program. From a practical standpoint, findings imply that cytostatic gene therapy using *RB* (or *RB* and *p53* together) may result in differential elimination of tumor cells through cellular senescence and crisis. At the same time the replicative lifespan of normal cells *in vivo* may not be affected. This could provide a potential basis for designing tumor-specific tumor suppressor gene therapy and anti-telomerase gene therapy.

These findings, taken together, may intimate that the *RB*-mediated tumor suppression has a broad biological basis, which certainly makes the emerging *RB* tumor suppressor gene therapy for human cancer even more attractive.

M. Enhanced Tumor Suppression by an N-terminal Truncated pRB.

Long-term stable clones of the *RB*-reconstituted tumor cells can be isolated with non-inducible gene expression systems, and most of these clones grow just as rapidly as the parental lines. The inventors have also found that, although the *RB*-mediated tumor suppression was substantial and had a broad biological basis, it was often incomplete and a portion of the *RB*-reconstituted tumor cells were able to survive and form RB^+ xenograft tumors in nude mice after a prolonged latency period (Takahashi *et al.*, 1991; Xu *et al.*, 1991b; Zhou *et al.*, 1994b; Li *et al.*, 1996). Similar observations have been reported by other investigators (Bookstein *et al.*, 1990b; Goodrich *et al.*, 1992b; Kratzke *et al.*, 1993; Ookawa *et al.*, 1993; Wang *et al.*, 1993). This phenomenon is referred to by the inventors as *tumor suppressor resistance* (TSR; Zhou *et al.*, 1994b), which is an equivalent of multiple drug resistance (MDR) in chemotherapeutics. In the latter scenario, low-dose chemotherapy may risk the selection of metastatic tumor cells due to their often inherently higher resistance to cytotoxic agents.

The inventors subsequently reported that an N-terminal truncated RB protein of ~94 kDa (pRB⁹⁴) exerted surprisingly more potent cell growth suppression as compared to the full-length pRB protein in a diversity of tumor cell lines examined, including those having a normal endogenous RB gene. Tumor cells transfected with the pRB⁹⁴-expressing plasmids displayed multiple morphological changes frequently associated with cellular senescence. They failed to enter S phase and rapidly died (Xu *et al.*, 1994b; Resnitzky and Reed, 1995).

The inventors recent studies in ectopic animal models demonstrated that treatment of established human RB⁻ and RB⁺ bladder xenograft cancers in nude mice by AdCMVpRB94, a replication-deficient adenovirus vector expressing the N-terminal truncated RB protein, resulted in regression of the treated tumors (Xu *et al.*, 1996). Of note, although both the full-length and the truncated forms of the RB protein, when over-expressed in tumor cells via adenovirus vectors, were capable of suppression of tumor growth, the pRB⁹⁴ was much more potent than the full-length RB protein. The mechanism for the enhanced tumor suppression by the N-terminal truncated RB protein is not clear yet.

To better understand the functional difference between the N-terminal truncated pRB⁹⁴ and the full-length pRB¹¹⁰, the inventors have also established stable tumor cell lines with Tc-responsive pRB⁹⁴ expression. By time course analysis, it was found that as early as 6 hours after removal of tetracycline from the cell culture medium, the pRB⁹⁴-reconstituted tumor cells accumulated the maximum of both underphosphorylated and phosphorylated pRB⁹⁴, followed by failure of the vast majority of the tumor cells to incorporate ³H-thymidine, an indicator of growth cessation. The pRB⁹⁴ protein was completely dephosphorylated within ~18 to 24 hours. Most of the pRB¹¹⁰-reconstituted tumor cells, however, remained immuno-histochemically RB⁻ at the 6 or 8 hr-time points and had normal DNA synthesis (FIG. 5). The pRB¹¹⁰ reached the highest level at the 24 hr-time point as determined by western blotting, and became mostly unphosphorylated from 24 to 48 hours after removal of tetracycline, in which period the pRB¹¹⁰-

reconstituted tumor cells finally ceased DNA synthesis (FIG. 5). Using the SA- β -gal biomarker assay for human senescent cells, it was shown that the Saos-2 cells with pRB⁹⁴ expression showed more intense SA- β -gal positive staining as compared to the pRB¹¹⁰-expressing cells at 48 hr after removal of Tc. Since pRB⁹⁴ has a longer half-life than pRB¹¹⁰ and tends to remain in an active, underphosphorylated form (U. S. Patent 5,496,731; Xu *et al.*, 1994b), rapid accumulation of mostly the active forms (underphosphorylated form) of RB protein in the tumor cells may account for the enhanced tumor cell growth suppression by pRB⁹⁴. In this regard, another truncated version of pRB, named pRB⁵⁶, beginning at amino acid 379, has also been reported as a more potent inhibitor of cell cycle progression compared to the full-length pRB (Wills *et al.*, 1995).

The advantages of the modified system are threefold: 1) it is suitable for establishing long-term stable cell lines with inducible gene expression because of lower constitutive expression of the tTA peptide; 2) the system is now contained within a single plasmid so that only one round of transfection and selection is required; and 3) of importance, the single-plasmid tetracycline-responsive mammalian gene expression system is readily convertible to tetracycline-controlled viral vectors (Examples 7-12 below).

EXAMPLE 7

Construction of Tetracycline-Controlled Adenoviral Vectors

The desired cDNA fragment of a gene of interest is first inserted into the single-plasmid tetracycline-regulatable plasmid vector, EC1214A (Example 3) or EC1214B (Example 4). The tetracycline-responsive foreign gene expression cassette and the modified tTA (or rtTA) expression cassette from the corresponding EC1214A or EC1214B plasmid vectors are then recovered using standard methods in the art for DNA manipulation (Maniatis *et al.*, 1989; Ausubel *et al.*, 1992), and inserted into the shuttle plasmid, p Δ E1sp1A (Microbix Biosystems, Inc.). The resultant recombinant shuttle plasmids are then co-transfected with the master adenovirus type 5 (Ad5) plasmid, pBHG11, which contains the backbone of the adenovirus Ad5dl309 genome and E1/E3 deletion mutation (Microbix Biosystems, Inc.) into 293 cells using

the LIPOFECTIN reagent (GIBCO/BRL Life Technologies). The co-transfection of 293 cells is performed in the presence (for tet-off system) or absence (for tet-on system) of 0.5 µg/ml of tetracycline.

Alternatively, a fragment containing a gene of interest is first inserted into the single-plasmid tetracycline-regulatable plasmid vector, EC1214A or EC1214B. The tetracycline-responsive foreign gene expression cassette and the modified tTA (or rtTA) expression cassette from the corresponding EC1214A or EC1214B plasmid vectors are then recovered and inserted, respectively, into the shuttle plasmid, pΔE1sp1A and the master adenovirus plasmid, pBHG11. The resultant recombinant shuttle plasmids and the recombinant master adenovirus plasmid are co-transfected into 293 cells.

Co-transfection of 293 cells with the recombinant shuttle plasmid and the recombinant master adenovirus plasmid produce infectious virions by *in vivo* recombination, in which the minigene cassette expressing the gene of interest and the modified tTA (or rtTA) expression cassette are replaced the ΔE1 region or ΔE1 and ΔE3 regions of the Ad5dl309 genome, respectively. Presence of recombinant adenoviruses in the transfected 293 cells is initially identified by cytopathic effect (CPE). Cell culture supernatants are collected from the transfected 293 cells in which CPE has occurred. Recombinant viruses are then isolated by screening adenovirus plaques from 293 cell monolayers after infection with the virus supernatants, and further characterized by restriction enzyme digestion mapping, PCR™, or by expression of the gene of interest in virus-infected host cells in a tetracycline-regulatable manner. The recombinant adenoviruses containing the desired foreign gene as well as the modified tTA (or rtTA) expression cassettes are subjected to at least three rounds of plaque purification.

High-titer stocks of the tetracycline-controlled recombinant adenoviruses are prepared by methods modified from Graham and Prevec, (1991). The CsCl ultracentrifugation-purified adenoviruses contain $\sim 10^{13}$ viral particles per ml as measured by OD at 260 nm ($1 \text{ OD}_{260} = 1 \times 10^{12}$ viral particles per ml). The concentrated viral suspension is desalted by gel filtration

through Sephadex G50 to generate a final purified virus stock about 10^{11} plaque-forming units (pfu) per ml in PBS.

EXAMPLE 8

Preparation of Tetracycline-Responsive RB Adenovirus Vector

A replication-deficient adenovirus vectors expressing N-terminal truncated pRB⁹⁴ protein (U. S. Patent No. 5,496,731) has been used in *in vivo* animal studies of human cancer gene therapy (Xu *et al.*, 1996). Unfortunately, the ratio of viral particles to plaque-forming units of the AdCMVpRB94 virus supernatants increased dramatically with passage, making it difficult for large-scale preparation of high-titer stocks of the AdCMVpRB94 virus for human cancer gene therapy clinical trials. This was probably caused by the super cell growth suppression effects of pRB94 protein on the 293 virus-producing cell line.

The modified tetracycline-responsive mammalian gene expression system has been used in a similar manner as described above to generate a tetracycline-controlled pRB⁹⁴-containing adenovirus vector, AdvtTA.RB94, which is designed for delivery of high-dose pRB⁹⁴ gene therapy. The entire tetracycline regulation cassette can be inserted into the E1 region of the adenovirus genome, or the RB⁹⁴ expression cassette can be inserted into the E1 region of the adenovirus genome, while the transcriptional transactivation fusion protein expression cassette is inserted into the E3 region of the adenovirus genome. Over-expression of pRB⁹⁴ in tumor cells will cause tumor cell-specific senescence and cell death. The pRB⁹⁴ cDNA has a modified optimal initiator context sequence. Expression of the pRB94 protein in transduced human tumor cells by AdvtTA.RB94 can be reversibly turned off and on. The novel AdvtTA.RB94 recombinant adenovirus vector can be propagated efficiently in 293 cells with increased yield and quality.

EXAMPLE 9

Preparation of Tetracycline-Responsive RB/p53 Coexpression Vector

As described in Example 6 above, although p53 reconstitution in Saos-2 (RB⁻, p53^{null}) tumor cells with either non-inducible (Chen *et al.*, 1990; Li *et al.*, 1996) or inducible system did suppress their neoplastic phenotype, the p53 reconstituted Saos-2 clones with the tetracycline-regulatable promoter were SA-β-gal negative in either presence or absence of tetracycline. However, when the p53-reconstituted Saos-2 cells were infected with recombinant adenovirus vectors expressing the wild-type pRB¹¹⁰ in Tc-free medium, the tumor cells with both wild-type p53 and pRB¹¹⁰ expression displayed more intense SA-β-gal positive staining as compared to tumor cells only expressing pRB¹¹⁰. The results imply that the mechanisms for tumor suppression by pRB and p53 were different from each other, but expression of pRB and p53 together had synergistic effects on RB-mediated tumor cell senescence.

Since co-expression of pRB and p53 has synergistic effects on pRB-mediated, tumor-specific senescence (Example 6), and it has been suggested that altered RB and p53 protein status could be a synergistic prognostic factor in non-small cell lung carcinomas, as well as a subset of other human malignancies, including transitional cell carcinomas of the bladder (Xu, 1995; Xu *et al.*, 1994a; Xu *et al.*, 1996), combination pRB and p53 gene therapy is also contemplated as an alternative strategy to surmount possible tumor suppressor resistance.

Insertion of both the modified tetracycline-responsive transactivator (tTA) expression cassette and the tTA-dependent pRB¹¹⁰ expression cassette into the E1 region of the Ad5 genome facilitates construction of an adenovirus vector simultaneously expressing two tumor suppressor genes, named AdvtTA.RB110/p53. In this vector, the smaller p53 expression cassette is inserted into the E3 region of the 34 kb master plasmid, pBHG11, through ligation reaction. Since attempts to replace both RB and p53 genes in the same cell have never been successful (Wang *et al.*, 1993), the inventors reasoned that adenovirus vectors simultaneously expressing the two tumor suppressor genes should be built in the regulatable gene expression system.

EXAMPLE 10

Construction of Tetracycline-Controlled Retroviral Vectors

The *kat* retrovirus production system produces high titer retrovirus supernatant capable of transducing efficiently hematopoietic cell types refractory to conventional retrovirus transduction (Finer *et al.*, 1994). The *kat* retrovirus plasmid vector with a hybrid LTR will be combined with EC1214A (Example 3) to generate a retrovirus with Tc-regulatable expression. Since some success using standard retroviral vectors have been reported in the literature, the Tc-controlled retroviral vector may work better than the Tc-controlled adenoviral vector for transduction of certain cell types, such as hematopoietic stem cells.

EXAMPLE 11

Therapeutic Administration of Modified RB Constructs

A. Treatment of Human Bladder Cancers *in vivo*.

The human bladder cancer represents an ideal model for practicing tumor suppressor gene therapy of solid tumors by infusing the instant modified RB protein expression retroviral vectors into the bladder. The original experimental model of human bladder cancer was established by Jones and colleagues (Ahlering *et al.*, 1987). It has been shown that human bladder tumor cells of RT4 cell line established from a superficial papillary tumor, which usually does not metastasize, produced tumors only locally when injected by a 22-gauge catheter into the bladder of female nude mice. In contrast, the EJ bladder carcinoma cells which were originally isolated from a more aggressive human bladder cancer produced invasive tumors in the nude mouse bladders which metastasized to the lung spontaneously. Therefore, this model can be used for treatment of experimental bladder cancer by *in vivo* gene transfer with retroviral vectors.

Tumor cells from RB minus human bladder carcinoma cell line, 5637 (ATCC HTB9) and RB⁺ human bladder carcinoma cell line, SCaBER (ATCC HTB3) will be injected directly into the bladders of female athymic (nu/nu) nude mice (6 to 8 weeks of age) by a catheter as initially reported by Jones and colleagues (Ahlering *et al.*, 1987). Development and progression of the nude mouse bladder tumors will be monitored using a fiber-optical system to which a TV

monitor is attached. The experimental tumors will subsequently be treated with retrovirus vectors expressing the modified RB proteins of the present invention.

Supernatants with high virus titers will be obtained from tissue culture media of selected clones expressing high level of human modified RB protein and confirmed as free of replication-competent virus prior to use. The retroviral vector suspension at high titers ranging from 4×10^4 to greater than 1×10^7 colony-forming unit (cfu)/ml, and more preferably at a titer greater than 1×10^6 cfu/ml will then be infused directly into the mouse bladders *via* a catheter to treat the tumors. The skilled artisan will understand that such treatments may be repeated as many times as necessary *via* a catheter inserted into the bladder. The tumor regression following transferring the modified RB gene will be monitored frequently *via* the fiber-optic system mentioned above.

The same procedure as described above may be used for treating the human bladder cancer except that the retroviral vector suspension is infused into a human bladder bearing cancer.

B. *in vivo* Studies Using an Orthotopic Lung Cancer Model

Human large cell lung carcinoma, NCI-H460 (ATCC HTB177) cells which have normal pRB¹¹⁰ expression will be injected into the right mainstream bronchus of athymic (nu/nu) nude mice (10^5 cells per mouse). Three days later the mice will be inoculated endobronchically with supernatant from the modified RB, or wild-type RB retrovirus producer cells daily for three consecutive days. Tumor formation suppression in the group of mice treated with the modified RB retrovirus supernatant, in contrast, to the group which is treated with wild-type RB retrovirus supernatant, will indicate that the modified RB-expressing retrovirus inhibits growth of RB⁺ non-small cell lung carcinoma (NSCLC) cells, whereas the wild-type RB-expressing retrovirus does not.

C. Treatment of Human Non-Small Cell Lung Cancers *in vivo*.

Non-small cell lung cancer patients having an endobronchial tumor accessible to a bronchoscope, and also having a bronchial obstruction, will be initially selected for modified RB

gene therapy. Treatment will be administered by bronchoscopy under topical or general anesthesia. To begin the procedure, as much gross tumor as possible will be resected endoscopically. A transbronchial aspiration needle (21G) will be passed through the biopsy channel of the bronchoscope. The residual tumor site will then be injected with the appropriate modified RB retroviral vector supernatant, modified RB adenovirus suspension or modified RB-expressing plasmid vector-liposome complexes at a volume of 5 ml to 10 ml. Protamine may be added to a concentration of 5 µg/ml. The injections of therapeutic viral or plasmid supernatant comprising one or more of the vectors will be administered around and within the tumor or tumors and into the submucosa adjacent to the tumor. The injections will be repeated daily for five consecutive days and monthly thereafter. The treatment may be continued as long as there is no tumor progression. After one year the patients will be evaluated to determine whether it is appropriate to continue therapy.

In addition, as a precaution, the patients will wear a surgical mask for 24 hours following injection of the viral supernatant. All medical personnel will wear masks routinely during bronchoscopy and injection of the viral supernatant. Anti-tussive will be prescribed as necessary.

D. Treatment or Prevention of Human Lung Carcinomas With Liposome-Encapsulated Purified Modified RB Protein

In yet another alternative, target tumor or cancer cells will be treated by introducing the instant modified RB proteins into cells in need of such treatment by any known method. For example, liposomes are artificial membrane vesicles that have been extensively studied for their usefulness as delivery vehicles of drugs, proteins and plasmid vectors both *in vitro* or *in vivo* (Mannino *et al.*, 1988). Proteins such as erythrocyte anion transporter (Newton *et al.*, 1988), superoxide dismutase and catalase (Tanswell *et al.*, 1990), and UV-DNA repair enzyme (Ceccoli *et al.*, 1989) have been encapsulated at high efficiency with liposome vesicles and delivered into mammalian cells *in vitro* or *in vivo*. Further, small-particle aerosols provide a method for the delivery of drugs for treatment of respiratory diseases. For example, it has been reported that drugs can be administered in small-particle aerosols by using liposomes as a vehicle.

Administered *via* aerosols, the drugs are deposited rather uniformly on the surface of the nasopharynx, the tracheobronchial tree and in the pulmonary area (Knight *et al.*, 1988).

To treat or prevent lung cancers, the therapeutic modified RB proteins will be purified, for example, from recombinant baculovirus AcMNPV-modified RB infected insect cells by immunoaffinity chromatography or any other convenient source. The modified RB protein will then be mixed with liposomes and incorporated into the liposome vesicles at high efficiency. The encapsulated modified RB will still be active. Since the aerosol delivery method is mild and well-tolerated by normal volunteers and patients, the modified RB-containing liposomes can be administered to treat patients suffering from lung cancers of any stage and/or to prevent lung cancers in high-risk population. The modified RB protein-containing liposomes may be administered by nasal inhalation or by a endotracheal tube *via* small-particle aerosols at a dose sufficient to suppress abnormal cell proliferation. Aerosolization treatments will be administered to a patient for 30 minutes, three times daily for two weeks, with repetition as needed. The modified RB protein will thereby be delivered throughout the respiratory tract and the pulmonary area. The treatment may be continued as long as necessary. After one year, the overall condition of the patient will be evaluated to determine if continued therapy is appropriate.

EXAMPLE 12

Induction of Senescence and Telomerase Inhibition by Reexpression of RB

Normal human diploid cells senesce *in vitro* and *in vivo* after a limited number of cell divisions. This process, known as cellular senescence, is an underlying cause of aging and a critical barrier for development of human cancers. This Example presents studies that demonstrate that reexpression of functional pRB alone in *RB/p53*-defective tumor cells *via* a modified tetracycline-regulated gene expression system resulted in a stable growth arrest at the G0/G1 phase of the cell cycle, preventing tumor cells from entering S phase in response to a variety of mitogenic stimuli. These cells displayed multiple morphological changes consistent with cellular senescence and expressed a senescence-associated β -galactosidase biomarker.

Additionally, telomerase activity, which is believed to be essential for an extended proliferative life-span of neoplastic cells, was abrogated or repressed in the tumor cell lines after induction of pRB (but not p53) expression. Strikingly, when returned to a non-permissive medium for pRB expression, the pRB-induced senescent tumor cells resumed DNA synthesis and attempted to divide. However, most cells died in the process, a phenomenon similar to postsenescent crisis of SV40 T-antigen-transformed human diploid fibroblasts in late passage. These observations provide direct evidence that overexpression of pRB alone in *RB/p53*-defective tumor cells is sufficient to reverse their immortality and cause a phenotype that is, by all generally accepted criteria, indistinguishable from replicative senescence. The results indicate that pRB may play a causal role in the intrinsic cellular senescence program.

A. Materials and Methods

Establishing tumor cell lines with Tc-regulatable pRB expression

The original multiple-plasmid tetracycline repressor/operator-based regulatory system was improved as described in detail above. All *RB*-reconstituted tumor cell lines used in this Example were subjected to at least two rounds of subcloning following the initial plasmid transfection and are considered pure clones. The homogeneity of these clones was verified by pRB nuclear staining. In addition, a panel assay (Zhou *et al.*, 1994) was used to ensure stable expression of the functional pRB under permissive conditions. The *RB*-reconstituted tumor cells were all *RB*⁻ in the presence of 0.5 µg/ml of Tc in culture medium; while the great majority (>99%) of the cells became *RB*⁺ at 24 hours after removal of Tc as shown by immunocytochemical staining.

Flow cytometric analysis

Single cell suspensions collected at each time point were fixed with paraformaldehyde and ethanol before propidium iodide (PI) (Sigma) staining. All profiles were generated using a FACScan flow cytometer (Becton-Dickinson). The first peak (M1) contains cells with diploid DNA in G0/G1, the second peak (M3) with twice the PI-fluorescence intensity contains tetraploid G2/M cells, and the area between the two peaks (M2) represents the total number of cells in S phase (Nicoletti *et al.*, 1991).

SA-β-gal assay

The assay was performed essentially as previously described (Dimri *et al.*, 1995). Briefly, the cells were fixed in 2% formaldehyde/0.2% glutaraldehyde for 5 min and stained with 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-Gal) at pH 6.0 for 6 hours. The staining solution contained 1 mg/ml X-Gal, 40 mM citric acid/sodium phosphate, pH 6.0, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 150 mM NaCl and 2 mM MgCl₂.

Telomeric repeat amplification protocol (TRAP) assay

The methodology, according to the technical manual, was modified from the original TRAP assay as described by Kim *et al.* (Kim *et al.*, 1994). In short, ~10⁶ cells grown in a 100-mm Petri dish were harvested and resuspended in 200 μl of ice-cold lysis buffer for 30 min on ice, followed by centrifugation at 100,000 × g for 30 min at 4°C. The supernatant was diluted to 0.5 μg protein/μl, of which 2 μl was used for each TRAP assay. The telomerase reaction was carried out at 30°C for 30 min, which was followed by a 2-step PCRTM amplification with [γ-³²P]-labeled TS primer (94°C, 30 s and 60°C, 30 s for 33 cycles). The PCRTM-amplified telomerase extension products were subjected to electrophoresis on a 12.5% polyacrylamide gel.

B. Results

pRB-mediated irreversible growth cessation of tumor cells

Using the modified tetracycline (Tc)-regulatable gene expression system as described in detail above, dozens of long-term stable tumor cell clones were established, in which expression of the wild-type pRB can be reversibly turned on and off without significant leakage. The *RB*-reconstituted tumor cell clones were obtained, respectively, from the breast carcinoma cell line, MDA-MB-468, the osteosarcoma cell line Saos-2, and the bladder carcinoma cell line, 5637. These tumor cell lines were chosen as host cells since they were known to contain both *RB* and *p53* gene mutations (Wang *et al.*, 1993; Chen *et al.*, 1990; Berry *et al.*, 1996; Masuda *et al.*, 1987).

As measured by western blotting, pRB protein induced in the tumor cells reached the highest level about 24 hours after removal of tetracycline from the cell culture medium, and then became completely dephosphorylated within 24 to 40 hours. The effects of induction of pRB expression on tumor cell growth were subsequently examined in representative clones by measuring growth curves and (³H) thymidine incorporation (Xu *et al.*, 1994b), and by flow cytometric analysis (Nicoletti *et al.*, 1991). Cell growth and DNA synthesis of all the long-term tumor cell clones studied ceased 24 to 48 hours after pRB expression was induced (FIG. 3A, FIG. 3B and FIG. 3C). The great majority of the tumor cells were arrested at G0/G1 phase of the cell cycle.

After a 4-day induction of pRB expression in Tc-free medium, the growth cessation of the tumor cells was irreversible by stimulation with a variety of mitogens, such as serum growth factors, phytohemagglutinin (PHA) and concanavalin A (Con A). This was determined by continuous flat growth curves as shown in FIG. 3A, FIG. 3B and FIG. 3C and failure of the tumor cells to incorporate (³H) thymidine in response to mitogenic stimulation. In the meantime, the tumor cells displayed striking morphological changes consistent with cellular senescence, including cell enlargement, flattening, and lower nucleocytoplasmic ratio than cycling cells.

Furthermore, as measured by DNA fragmentation assay, a small amount of lower molecular weight DNAs were often observed in DNA samples prepared from *RB*-reconstituted Saos-2 tumor cells grown in non-permissive but not permissive conditions for pRB expression. This finding suggested a low level of spontaneous apoptosis of the *RB*-defective tumor cell culture, which was inhibited by induction of pRB expression. In addition, switching on pRB expression in the *RB*-reconstituted 5637 and MDA-MB-468 tumor cell lines also inhibited IFN- γ -induced apoptotic cell death.

Expression of senescence-associated β -galactosidase

A biomarker that identifies senescent human cells in culture and in aging skin *in vivo* has recently been reported. This marker, termed senescence-associated β -galactosidase (SA- β -gal), is expressed by senescent, but not pre-senescent fibroblasts. SA- β -gal was also absent from

immortal cells but was induced by genetic manipulations that reversed immortality (Dimri *et al.*, 1995). Young (early passage) human WI-38 fibroblasts were SA- β -gal negative, whereas the senescent (at population doubling level greater than 52) WI-38 cells were strongly SA- β -gal positive, which provided a valid control for the SA- β -gal assay. The Tc-responsive *RB*-reconstituted tumor cell clones were totally SA- β -gal negative in the presence of Tc (*i.e.*, in *RB*⁻ status), and the majority of the tumor cells became SA- β -gal positive after induction of p*RB* expression for four to five days in Tc-free medium. The detection of this senescence-associated biomarker in the tumor cells was coincident with the irreversible growth cessation of the tumor cell populations (FIG. 3A, FIG. 3B and FIG. 3C). The intensity of the SA- β -gal staining of the induced *RB*⁺ tumor cells, however, was variable depending on the tumor cell types.

*Reexpression of p*RB* (but not p53) in tumor cells inhibited telomerase activity*

Since telomerase has recently emerged as an attractive candidate for a regulator in cellular senescence (Linskens *et al.*, 1995; Klingelhutz *et al.*, 1996), the effects of p*RB* and p53 replacement on the telomerase activity of the host tumor cells were determined. In this connection, several long-term stable tumor cell clones with Tc-regulatable wild-type p53 expression from the osteosarcoma cell line, Saos-2 were established. A telomeric repeat amplification protocol (TRAP) assay as recently described (Kim *et al.*, 1994) was used to measure telomerase activity in tumor cells before and after induction of p*RB* (or p53) expression.

Prior to induction of p*RB* expression, the *RB*-reconstituted tumor cell clones from all three *RB*/p53-defective tumor types examined were positive for telomerase activity, whereas the relative telomerase activity was ~15 to >100 times lower in the tumor cells after turning on the p*RB* expression as estimated by densitometry of the digitized image. In fact, the telomerase activity was nearly non-detectable in the p*RB*-expressing MDA-MB-468 and Saos-2 tumor cells. In contrast, although induction of wild-type p53 expression in Saos-2 did result in growth arrest of the *RB*⁻/p53^{null} tumor cells, the p53-reconstituted Saos-2 tumor clones persistently exhibited positive telomerase activity, which was not affected by their p53 status. Thus the differences in telomerase activity cannot be explained simply as a difference in cell proliferation.

Postsenescent crisis of pRB-induced senescent tumor cells after withdrawal of pRB

The pRB-induced tumor cell senescence was stringently dependent on the continued expression of the functional pRB. As shown above, after induction of pRB expression in Tc-free medium for four or more days, the *RB*-reconstituted MDA-MB-468, Saos-2, and 5637 tumor cells became senescent. When these tumor cells returned to an non-permissive medium for pRB expression, however, a large number of tumor cells were observed that lost cell-cell adherence, detached from the Petri dishes and died. To further characterize this phenomenon, a combined method was employed involving pRB immunocytochemical staining and (³H) thymidine *in situ* labeling of the tumor cells.

It was found that after adding 0.5 µg/ml of Tc back to the *RB*-reconstituted Saos-2 tumor cell cultures that had been maintained in Tc-free medium for 4 to 5 days, nearly all tumor cells were depleted of the exogenous pRB and became RB⁻ at day 6. Subsequently, at day 9 to 10, the tumor cells resumed DNA synthesis, the majority of which however had strikingly aberrant nuclei. They attempted to divide but most died in the process. These tumor cells displayed a phenotype, showing remarkable similarity to postsenescent crisis of the T-antigen-transformed human cells in late passage (Stein, 1985).

In summary, reexpression of functional pRB in *RB*-defective tumor cells induced growth cessation concurrently with inhibition of telomerase activity. The tumor cells irreversibly lost mitogen responsiveness, entering a viable G1-arrested state. They also exhibited pRB-dependent SA-β-gal positivity (a senescence-associated biomarker) and resistance to apoptotic cell death. Of note, replacement of either wild-type pRB or p53 in the RB⁻/p53^{null} Saos-2 was able to block tumor cell growth at the population level, but only pRB induced inhibition of telomerase. Furthermore, withdrawal of pRB in pRB-induced senescent tumor cells led to a crisis-like phenotype. These observations, taken together, suggest pRB is causally involved in the cellular senescence program. These results are the first direct evidence that overexpression of pRB alone in a variety of *RB*-defective tumor cells was sufficient to reverse their immortality and cause *bona fide* replicative senescence. Since all three *RB*-defective tumor cell lines examined also

have p53 mutations, the pRB-mediated tumor cell senescence apparently do not require wild-type p53 function.

Thus a new link between pRB and telomerase is shown. It is demonstrated, by a telomeric repeat amplification protocol (TRAP) assay, that reexpression of pRB in *RB*-defective tumor cells inhibits telomerase activity. Because of the high sensitivity of the polymerase chain reaction (PCR™)-based TRAP assay, which detects the enzyme activity in a very small number of telomerase positive cells, and the difficulty in obtaining absolutely pure *RB*-reconstituted cell clones, the effectiveness of pRB reexpression on inhibition of telomerase activity in *RB*-defective tumor cells was likely even greater than it had been detected by the *in vitro* assay.

It is also noteworthy that, when maintained in non-permissive conditions for pRB (or p53) expression, the pRB-reconstituted Saos-2 clone apparently had much lower telomerase activity than the p53-reconstituted Saos-2 clone. The difference implies that, even before switching-on of the pRB expression in Tc-free medium, there must be low baseline expression of pRB from the Tc-responsive promoter in Saos-2 cells (Gossen and Bujard, 1995). The leakiness of pRB in pRB-reconstituted tumor cells under non-permissive conditions is below the immunodetection threshold for pRB protein (Xu *et al.*, 1991b), but it might be sufficient to inhibit the most telomerase activity. Since the tumor cells lacking telomerase activity likely resume telomere decline, this would eventually trigger the intrinsic cellular senescence program if it remains intact in the tumor cells.

* *

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and methods and in the steps or in the sequence of steps of the methods described herein without departing from the concept, spirit and scope of the invention. More specifically, it

will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

5

U. S. Patent 4,237,224

U. S. Patent 4,321,883

U. S. Patent 4,430,434

U. S. Patent 4,554,101

10

U. S. Patent 4,559,302

U. S. Patent 4,727,028

U. S. Patent 4,745,051

U. S. Patent 4,797,368

U. S. Patent 4,879,236

15

U. S. Patent 4,960,704

U. S. Patent 5,077,214

U. S. Patent 5,139,941

U. S. Patent 5,155,037

U. S. Patent 5,162,222

20

U. S. Patent 5,169,784

U. S. Patent 5,278,050

U. S. Patent 5,464,758

U. S. Patent 5,496,731

EPO 0273085

25

WO 90/05180

WO 91/15580

Ahlering *et al.*, *Cancer Res.*, 47:6660-6665, 1987.

Alt *et al.*, *J. Biol. Chem.*, 253:1357, 1978

Anderson *et al.*, *Genes Chromosom. Cancer*, 9:251-260, 1994.

5 Ausubel *et al.*, In: *Short Protocols in Molecular Biology, Current Protocols in Molecular Biology*, John Wiley and Jones, New York, 1992.

Bahnemann *et al.*, *Abs. Pap. ACS*, 180:5, 1980.

Baichwal and Sugden, In: *Gene Transfer*. R. Kucherlapati (ed.), New York, Plenum Press, pp. 117-148, 1986.

10 Baker *et al.*, *Science*, 249:912-915, 1990.

Ballester *et al.*, *Cell*, 63:851-859, 1990.

Benedict *et al.*, *J. Clin. Invest.*, 85:988-993, 1990.

Berry *et al.*, *Oncogene*, 12:1809-1819, 1996.

Bookstein *et al.*, *Crit. Rev. Oncog.*, 2(3):211-227, 1991.

15 Bookstein *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:7762-7766, 1990a.

Bookstein *et al.*, *Science*, 247:712-715, 1990b.

Boshart, *et al.*, *Cell*, 41: 521-530, 1985.

Brouckaert *et al.*, *Lymph. Cyt. Res.*, 11:193-196, 1992.

Brough *et al.*, *J. Virol.*, 70:6497-6501, 1996.

20 Buchkovich *et al.*, *Cell*, 58:1097-1105, 1989.

Burns *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:8033-8037, 1993.

Call *et al.*, *Cell*, 60:509-520, 1990.

Cance *et al.*, *New Engl. J. Med.*, 323:1457-1462, 1990.

Capaldi *et al.*, *Biochem. Biophys. Res. Comm.*, 76:425, 1977.

25 Cawthon *et al.*, *Cell*, 62:193-201, 1990.

Ceccoli *et al.*, *Journal of Investigative Dermatology*, 93:190-194, 1989.

Chalfie *et al.*, *Science*, 263:802-805, 1994.

Chang *et al.*, *Science*, 267:518-522, 1995.

Chang *et al.*, *Hepatology*, 14:134A, 1991.

- Chellappan *et al.*, *Cell*, 65:1053-1061, 1991.
- Chen and Okayama, *Mol. Cell. Biol.*, 7:2745-2752, 1987.
- Chen *et al.*, *Cell*, 58:1193-1198, 1989.
- Chen *et al.*, *Cell Growth Differ.*, 3:119-125, 1992.
- 5 Chen *et al.*, *Science*, 250:1576-1580, 1990.
- Choi *et al.*, *Mol. Cell. Biol.*, 11:3070-4, 1991.
- Clark *et al.*, *Human Gene Therapy*, 6:1329-1341, 1995.
- Coffin, In: *Virology*, Fields *et al.* (eds.), New York, Raven Press, pp. 1437-1500, 1990.
- Colberre-Garapin *et al.*, *J. Mol. Biol.*, 150:1, 1981.
- 10 Cooper and Whyte, *Cell*, 58:1009-1011, 1989.
- Cordon-Cardo *et al.*, *J. Natl. Cancer Inst.*, 84:1251-1256, 1992.
- Cotten *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:6094-6098, 1992.
- Couch *et al.*, *Am. Rev. Resp. Dis.*, 88:394-403, 1963.
- Coupar *et al.*, *Gene*, 68:1-10, 1988.
- 15 Curiel, In: *Viruses in Human Gene Therapy*, Vos (Ed.), Carolina Academic Press, Durham, pp. 179-212, 1994.
- Dawson *et al.*, *Proc. Am. Assoc. Cancer Res.*, 36:88, Abstract, 1995.
- Dawson *et al.*, Special AACR Conference: Cancer Susceptibility Genes and Molecular Carcinogenesis, Keystone, CO, February 19-25, 1996.
- 20 DeCaprio *et al.*, *Cell*, 58:1085-1095, 1989.
- Diller *et al.*, *Mol. Cell Biol.*, 10(11):5772-5781, 1990.
- Dimri *et al.*, *Proc. Natl. Acad. Sci. USA*, 92:9363-9367, 1995.
- Dowling, In: *Antibiotics Monographs*, No. 3, Medical Encyclopedia, New York, 1955.
- Duan *et al.*, *Science*, 269:1402-1406, 1995.
- 25 Evans, In: *Biochemical Reference Series 1*, Quadrangle Press, New York, 1968.
- Ewen *et al.*, *Cell*, 73:487-497, 1993.
- Fearon *et al.*, *Cell*, 61:759-767, 1990a.
- Fearon *et al.*, *Science*, 247:49-56, 1990b.
- Fechheimer *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:8463-8467, 1987
- 30 Finer *et al.*, *Blood*, 83:43-50, 1994.

- Finlay *et al.*, *Cell*, 57:1083-1093, 1989.
- Flotte *et al.*, *Gene Therapy*, 2:29-37, 1995.
- Flotte *et al.*, *Am. J. Respir. Cell Mol. Biol.*, 7:349-356, 1992.
- Flotte *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:10613-10617, 1993.
- 5 Fraley *et al.*, *Proc. Natl. Acad. Sci. USA*, 76:3348-3352, 1979
- Friedmann, *Science*, 244:1275-1281, 1989.
- Friend *et al.*, *Nature*, 323:643-646, 1986.
- Friend *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:9059-9063, 1987.
- Fung *et al.*, *Science*, 236:1657-1660, 1987.
- 10 Fung *et al.*, *Oncogene*, 8:2659-2672, 1993.
- Furukawa *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:2770-2774, 1990.
- Futreal *et al.*, *Science*, 266:120-122, 1994.
- Geng *et al.*, *Oncogene*, 12:1173-1180, 1996.
- Gessler *et al.*, *Nature*, 343:774-778, 1990.
- 15 Ghosh-Choudhury *et al.*, *EMBO J.*, 6:1733-1739, 1987.
- Ghosh and Bachhawat, *In: Liver diseases, targeted diagnosis and therapy using specific receptors and ligands*. Wu and Wu (eds.), New York, Marcel Dekker, pp. 87-104, 1991.
- Gill and Ptashne, *Nature*, 334:721-724, 1988.
- Gillies *et al.*, *Analytical Biochemistry* 159:109-113, 1986.
- 20 Gomez-Foix *et al.*, *J. Biol. Chem.*, 267:25129-25134, 1992.
- Goodrich and Lee, *Biochimica et Biophysica Acta*, 1155:43-61, 1993.
- Goodrich *et al.*, *Nature*, 360:177-179, 1992a.
- Goodrich *et al.*, *Cancer Res.*, 52:1968-1973, 1992b.
- Gopal, *Mol. Cell. Biol.*, 5:1188-1190, 1985.
- 25 Gossen and Bujard, *Proc. Natl. Acad. Sci. U. S. A.* 89:5547-5551, 1992.
- Gossen M and Bujard, *Biotechniques*, 19:213-216, 1995.
- Gossen *et al.*, *Science*, 268: 1766-1769, 1995.
- Graham and Prevec, *Biotechnology*, 20:363-390, 1992.
- Graham and Prevec, *In: Methods in Molecular Biology, Vol. 7: Gene Transfer and Expression*
- 30 *Protocols*, Murray (ed.), Clifton: the Humana Press Inc., pp. 109-128, 1991.

- Graham and Van Der Eb, *Virology* 52:456-467, 1973.
- Graham *et al.*, *J. Gen. Virol.*, 36:59-72, 1977.
- Grunhaus and Horwitz, *Seminar in Virology*, 3:237-252, 1992.
- Hamel *et al.*, *FASEB J.* 7:846-854, 1993.
- 5 Harland and Weintraub, *J. Cell Biol.*, 101:1094-1099, 1985.
- Hermonat and Muzyczka, *Proc. Natl. Acad. Sci. USA*, 81:6466-6470, 1984.
- Hersdorffer *et al.*, *DNA Cell Biol.*, 9:713-723, 1990.
- Herz and Gerard, *Proc. Natl. Acad. Sci. USA*, 90:2812-2816, 1993.
- Hlavka and Boothe, *In: Handbook of Experimental Pharmacology*, 78, Blackwood *et al.* (eds.),
 10 Springer-Verlag, Berlin-New York, 1985.
- Horowitz, *Genes Chromosom. Cancer*, 6:124-131, 1993.
- Horwich *et al.* *J. Virol.*, 64:642-650, 1990.
- Huang *et al.*, *Science*, 242:1563-1566, 1988.
- Jones and Shenk, *Cell*, 13:181-188, 1978.
- 15 Kamb *et al.*, *Science*, 264:436-440, 1994.
- Kaneda *et al.*, *Science*, 243:375-378, 1989.
- Kaplitt *et al.*, *Nature Genetics*, 8:148-154, 1994.
- Karlsson *et al.*, *EMBO J.*, 5:2377-2385, 1986.
- Kato *et al.*, *Genes Dev.*, 7:331-342, 1993.
- 20 Kato *et al.*, *J. Biol. Chem.*, 266:3361-3364, 1991.
- Kaufman, *Methods in Enzymology*, 185:537-566, 1990.
- Kelleher and Vos, *Biotechniques*, 17(6):1110-1117, 1994.
- Kim *et al.*, *Science*, 266:2011-2015, 1994.
- Klein *et al.*, *Nature*, 327:70-73, 1987.
- 25 Klein, *Bioessays*, 12:347-350, 1990.
- Klingelhutz *et al.*, *Nature*, 380:79-82, 1996.
- Knight *et al.*, *Eur. J. Clin. Micro. Infect. Dis.*, 7(6):721-731, 1988.
- Knudson, *Proc. Natl. Acad. Sci. USA*, 90:10914-10921, 1993.
- Knudson, *Cancer Res.*, 45:1437-1443, 1985.
- 30 Knudson, *Proc. Natl. Acad. Sci. USA*, 68:820-823, 1971.

- Koff *et al.*, *Science*, 257:1689-1694, 1992.
- Kornblau *et al.*, *Blood*, 84:256-261, 1994.
- Kotin *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:2211-2215, 1990.
- Kozak, *Nucl. Acids Res.*, 15(20):8125-8148, 1987.
- 5 Kozak, *J. Biol. Chem.*, 266:19867-19870, 1991.
- Kratzke *et al.*, *Cell Growth Differ.*, 4:629-635, 1993.
- Kunkel *et al.*, *Methods Enzymol.*, 154:367-382, 1987.
- Kyte and Doolittle, *J. Mol. Biol.*, 157:105-132, 1982.
- La Thangue, *Trends in Biochemical Sciences*, 19:108-114, 1994.
- 10 LaFace, *Viology*, 162:483-486, 1988.
- Larsson and Litwin, *Dev. Biol. Standard.*, 66:385-390, 1987.
- Laughlin *et al.*, *J. Virol.*, 60:515-524, 1986.
- Le Gal La Salle *et al.*, *Science*, 259:988-990, 1993.
- Lebkowski *et al.*, *Mol. Cell. Biol.*, 8:3988-3996, 1988.
- 15 Lee *et al.*, *Nature*, 329:642-645, 1987b.
- Lee *et al.*, *Science*, 235:1394-1399, 1987a.
- Levrero *et al.*, *Gene*, 101:195-202, 1991.
- Li *et al.*, *Oncogene*, 13:2379-2386, 1996.
- Lieber *et al.*, *J. Virol.* 70:8944-8960, 1996.
- 20 Linskens *et al.*, *Science*, 267:17, 1995.
- Logothetis *et al.*, *J. Natl. Cancer Inst.*, 84:1256-1261, 1992.
- Lu *et al.*, *Journal of Immunology*, 156:2495-2502, 1996.
- Lu *et al.*, *Oncogene*, 9:1015-1019, 1994.
- Ludlow *et al.*, *Cell*, 56:57-65, 1989.
- 25 Ludlow *et al.*, *Cell*, 60:387-396, 1990.
- Luo *et al.*, *Blood*, 82(Supp.):1,303A, 1994.
- Madreperla *et al.*, *Cancer Res.*, 51:6381-6384, 1991.
- Maniatis *et al.*, "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989.
- 30 Mann *et al.*, *Cell*, 33:153-159, 1983.

- Mannino *et al.*, *Biotechniques*, 6(7):682-690, 1988.
- Markowitz *et al.*, *J. Virol.*, 62:1120-1124, 1988.
- Marshall, *Cell*, 64:313-326, 1991.
- Masuda *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:7716-7719, 1987.
- 5 McCarty *et al.*, *J. Virol.*, 65:2936-2945, 1991.
- McGee *et al.*, *Gene*, 80:119-128, 1989.
- McLaughlin *et al.*, *J. Virol.*, 62:1963-1973, 1988.
- Michael, *Biotechniques*, 16:410-412, 1994.
- Mihara *et al.*, *Science*, 246:1300-1303, 1989.
- 10 Miki *et al.*, *Science*, 266:66-71, 1994.
- Mitschef, In: *Medicinal Research*, 9, Dekker, New York, 1978.
- Mittnacht *et al.*, *EMBO J.*, 13:118-127, 1994.
- Mizrahi, *Process Biochem.*, (August):9-12, 1983.
- Muncaster *et al.*, *Cancer Res.*, 52:654-661, 1992.
- 15 Murphree and Benedict, *Science*, 223:1028-1033, 1984.
- Muzyczka, *Curr. Top. Microbiol. Immunol.*, 158:97-129, 1992.
- Nevins, *Science*, 258:424-429, 1992.
- Newton *et al.*, *Biochemistry*, 27:4655-4659, 1988.
- Nicolas and Rubinstein, In: *Vectors: A survey of molecular cloning vectors and their uses*,
 20 Rodriguez and Denhardt (eds.), Stoneham: Butterworth, pp. 494-513, 1988.
- Nicolau and Sene, *Biochim. Biophys. Acta*, 721:185-190, 1982.
- Nicolau *et al.*, *Methods Enzymol.*, 149:157-176, 1987.
- Nicoletti *et al.*, *J. Immunol. Methods*, 139:271-279, 1991.
- Nilsson and Mosbach, *Dev. Biol. Standard.*, 66:183-193, 1987.
- 25 Noye Development Corporation, In: *Chemical Process Reviews*, Park Ridge, 2 volumes, 1969.
- O'Reilly *et al.*, In *Baculovirus expression vectors*, W.H. Freeman and Company, N.Y, 1992.
- Ohi *et al.*, *Gene*, 89L:279-282, 1990.
- Ookawa *et al.*, *Oncogene*, 8:2175-2181, 1993.
- Orita *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:2766-2770, 1989.
- 30 Pardee, *Science*, 246:603-608, 1989.

- Paskind *et al.*, *Virology*, 67:242-248, 1975.
- Perales *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:4086-4090, 1994.
- Petricciani, *Dev. Biol. Standard.*, 66:3-12, 1985.
- Phillips *et al.*, In: *Large Scale Mammalian Cell Culture*, Feder and Tolbert (eds.), Academic Press, Orlando, FL, 1985.
- Potter *et al.*, *Proc. Natl. Acad. Sci. USA*, 81:7161-7165, 1984.
- Pritchard-Jones *et al.*, *Nature*, 346:194-197, 1990.
- Qin *et al.*, *Genes Dev.*, 6:953- 964, 1992.
- Racher *et al.*, *Biotechnology Techniques*, 9:169-174, 1995.
- Ragot *et al.*, *Nature*, 361:647-650, 1993.
- Renan, *Radiother. Oncol.*, 19:197-218, 1990.
- Resnitzky and Reed, *Mol. Cell Biol.*, 15:3463-3469, 1995.
- Rich *et al.*, *Hum. Gene Ther.*, 4:461-476, 1993.
- Ridgeway, In: *Vectors: A survey of molecular cloning vectors and their uses*. Rodriguez and Denhardt (eds.) Stoneham: Butterworth, pp. 467-492, 1988.
- Riley *et al.*, *Annual Review of Cell Biology*, 10:1-29, 1994.
- Rippe *et al.*, *Mol. Cell Biol.*, 10:689-695, 1990.
- Rosenfeld *et al.*, *Science*, 252:431-434, 1991.
- Rosenfeld *et al.*, *Cell*, 68:143-155, 1992.
- Roux *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:9079-9083, 1989.
- Samulski *et al.*, *J. Virol.*, 63:3822-3828, 1989.
- Samulski *et al.*, *EMBO J.*, 10:3941-3950, 1991.
- Santerre, *et al.*, *Gene*, 30:147, 1984.
- Sardet *et al.*, *Proc. Natl. Acad. Sci. USA*, 92:2403-2407, 1995.
- Schott *et al.*, *Cancer Res.*, 54:1393-1396, 1994.
- Serrano *et al.*, *Nature*, 366:704-707, 1993.
- Shan *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:679-684, 1996.
- Shan *et al.*, *Mol. Cell Biol.*, 14:299-309, 1994.
- Shelling and Smith, *Gene Therapy*, 1:165-169, 1994.
- Stein *et al.*, *Science*, 249:666-669, 1990.

- Stein, *J. Cell Physiol.*, 125:36-44, 1985.
- Stratford-Perricaudet and Perricaudetp. In: *Human Gene Transfer*, Cohen-Haguenauer and Boiron, (eds.), Editions John Libbey Eurotext, France, P 51-61, 1991.
- Stratford-Perricaudet et al., *Hum. Gene. Ther.*, 1:241-256, 1991.
- 5 Sumegi et al., *Cell Growth Differ.*, 1:247-250, 1990.
- Summers and Smith, *Tex. Agric. Exp. Stn. Bull.* No.1555, 1987.
- Takahashi et al., *Proc. Natl. Acad. Sci. USA*, 88:5257-5261, 1991.
- Tanswell et al., *Biochemica et Biophysica Acta*, 1044:269-274, 1990.
- Temin, In: *Gene Transfer*, Kucherlapati (ed.), New York: Plenum Press, pp. 149-188, 1986.
- 10 Templeton et al., *Proc. Natl. Acad. Sci. USA*, 88:3033-3037, 1991.
- Tomic et al., *Nucl. Acids Res.*, 12:1656, 1990.
- Top et al., *J. Infect. Dis.*, 124:155-160, 1971.
- Tratschin et al., *Mol. Cell. Biol.*, 4:2072-2081, 1984.
- Tratschin et al., *Mol. Cell. Biol.*, 5:32581-3260, 1985.
- 15 Tur-Kaspa et al., *Mol. Cell Biol.*, 6:716-718, 1986.
- Upender et al., *Biotechniques*, 18:29-31, 1995.
- van Wezel, *Nature*, 216:64-65, 1967.
- Viskochil et al., *Cell*, 62:187-192, 1990.
- Wagner et al., *Science*, 260:1510-1513, 1990.
- 20 Wallace et al., *Science*, 249:181-186, 1990.
- Walsh et al., *J. Clin. Invest.*, 94:1440-1448, 1994.
- Wang et al., *Advances in Cancer Res.*, 64:25-85, 1994.
- Wang et al., *Oncogene*, 8:279-288, 1993.
- Watson et al., *Molecular Biology of the Gene*, 4th Ed., Benjamin, Inc., Menlo Park, CA, 1987.
- 25 Wei et al., *Gene Therapy*, 1:261-268, 1994.
- Weinberg, *Cancer Res.*, 49:3713-3721, 1989.
- Weinberg, *Cell*, 81:323-330, 1995.
- Weinberg, *Science*, 254:1138-1146, 1991.
- Weintraub et al., *Nature*, 358:259-261, 1992.
- 30 Weiss et al., *RNA Tumor Viruses*, Cold Spring Harbor, New York, 1985.

- Whyte *et al.*, *Nature*, 334:124-129, 1988.
- Wickham *et al.*, *Nature Biotechnology* 14:1570-1573, 1996.
- Wills *et al.*, *Cancer Gene Therapy*, Abstract, 2:339. 1995.
- Wong *et al.*, *Gene*, 10:87-94, 1980.
- 5 Wu and Wu, *J. Biol. Chem.*, 262:4429-4432, 1987.
- Wu and Wu, *Adv. Drug Delivery Rev.*, 12:159-167, 1993.
- Xu *et al.*, *Cancer Res.*, 51:2735-2739, 1991a.
- Xu *et al.*, *J. Natl. Cancer Inst.*, 86:695-699, 1994a.
- Xu *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:9837-9841, 1994b.
- 10 Xu *et al.*, *Cancer Res.*, 51:4481-4485, 1991c.
- Xu *et al.*, *Oncogene*, 6:1139-1146, 1991b.
- Xu *et al.*, *Oncogene*, 4:807-812, 1989b.
- Xu *et al.*, *Virology*, 171:331-341, 1989a.
- Xu, *Advances in Anatomic Pathology*, 2:213-226, 1995.
- 15 Xu, in *Advances in Pharmacology, Gene Therapy, Volume I*, Thomas August, ed., 1996.
- Xu, *et al.*, *Cancer Res. (Advances in Brief)*, 56:2245-2249, 1996.
- Yang *et al.*, *J. Virol.*, 68:4847-4856, 1994.
- Yang *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:9568-9572, 1990.
- Yoder *et al.*, *Blood*, 82 (Supp.): 1:347A, 1994.
- 20 Yokota *et al.*, *Oncogene*, 3:471-475, 1988.
- Zhou *et al.*, *Exp. Hematol. (NY)*, 21:928-933, 1993.
- Zhou *et al.*, *J. Exp. Med.*, 179:1867-1875, 1994a.
- Zhou *et al.*, *Proc. Natl. Acad. Sci. U S A*, 91:4165-4169, 1994b.
- Zou *et al.*, *Science*, 263:526-529, 1994.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Xu, Hong-Ji
Hu, Shi-Xue
Benedict, William F.
Zhou, Yunli

(ii) TITLE OF INVENTION: MODIFIED RETINOBLASTOMA TUMOR SUPPRESSOR
PROTEINS

(iii) NUMBER OF SEQUENCES: 51

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Arnold, White & Durkee
(B) STREET: P.O. Box 4433
(C) CITY: Houston
(D) STATE: TX
(E) COUNTRY: USA
(F) ZIP: 77210-4433

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US UNKNOWN
(B) FILING DATE: Concurrently Herewith
(C) CLASSIFICATION: UNKNOWN

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/038,118
(B) FILING DATE: 20-FEB-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Hibler, David W.
(B) REGISTRATION NUMBER: 41,071
(C) REFERENCE/DOCKET NUMBER: UTXC:506

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 512/418-3000
(B) TELEFAX: 512/474-7577

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3555 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 7..2790

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	384
Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
115 120 125	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	432
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
130 135 140	
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG	480
Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu	
145 150 155	
TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG	528
Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu	
160 165 170	

ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT GCA TTG GTG	576
Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val	
175 180 185 190	
CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA	624
Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu	
195 200 205	
CAA ATG GAA GAT GAT CTG GTG ATT TCA TTT CAG TTA ATG CTA TGT GTC	672
Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val	
210 215 220	
CTT GAC TAT TTT ATT AAA CTC TCA CCT CCC ATG TTG CTC AAA GAA CCA	720
Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro	
225 230 235	
TAT AAA ACA GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG	768
Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg	
240 245 250	
CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT	816
Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp	
255 260 265 270	
ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT	864
Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp	
275 280 285	
GAG GTG AAA AAT GTT TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT	912
Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser	
290 295 300	
CTT GGA CTT GTA ACA TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT	960
Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser	
305 310 315	
AAA CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA	1008
Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg	
320 325 330	
TTA TTT TTG GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT	1056
Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser	
335 340 345 350	
TTT GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG	1104
Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val	
355 360 365	
AAT GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC	1152
Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile	
370 375 380	

CAA CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA	1200
Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu	
385 390 395	
AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT	1248
Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser	
400 405 410	
ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT	1296
Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe	
415 420 425 430	
GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC	1344
Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr	
435 440 445	
AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA	1392
Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys	
450 455 460	
TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT	1440
Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn	
465 470 475	
GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA	1488
Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val	
480 485 490	
ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA	1536
Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr	
495 500 505 510	
GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT	1584
Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe	
515 520 525	
GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG	1632
Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu	
530 535 540	
ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG	1680
Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met	
545 550 555	
GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA	1728
Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys	
560 565 570	
CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT	1776
Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys	
575 580 585 590	

CCT	CTT	AAT	CTT	CCT	CTC	CAG	AAT	AAT	CAC	ACT	GCA	GCA	GAT	ATG	TAT	1824
Pro	Leu	Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	
				595					600					605		
CTT	TCT	CCT	GTA	AGA	TCT	CCA	AAG	AAA	AAA	GGT	TCA	ACT	ACG	CGT	GTA	1872
Leu	Ser	Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	
			610					615						620		
AAT	TCT	ACT	GCA	AAT	GCA	GAG	ACA	CAA	GCA	ACC	TCA	GCC	TTC	CAG	ACC	1920
Asn	Ser	Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	
		625					630					635				
CAG	AAG	CCA	TTG	AAA	TCT	ACC	TCT	CTT	TCA	CTG	TTT	TAT	AAA	AAA	GTG	1968
Gln	Lys	Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	
	640					645					650					
TAT	CGG	CTA	GCC	TAT	CTC	CGG	CTA	AAT	ACA	CTT	TGT	GAA	CGC	CTT	CTG	2016
Tyr	Arg	Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	
655					660					665					670	
TCT	GAG	CAC	CCA	GAA	TTA	GAA	CAT	ATC	ATC	TGG	ACC	CTT	TTC	CAG	CAC	2064
Ser	Glu	His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	
				675					680					685		
ACC	CTG	CAG	AAT	GAG	TAT	GAA	CTC	ATG	AGA	GAC	AGG	CAT	TTG	GAC	CAA	2112
Thr	Leu	Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	
			690					695					700			
ATT	ATG	ATG	TGT	TCC	ATG	TAT	GGC	ATA	TGC	AAA	GTG	AAG	AAT	ATA	GAC	2160
Ile	Met	Met	Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	
		705					710					715				
CTT	AAA	TTC	AAA	ATC	ATT	GTA	ACA	GCA	TAC	AAG	GAT	CTT	CCT	CAT	GCT	2208
Leu	Lys	Phe	Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	
	720					725					730					
GTT	CAG	GAG	ACA	TTC	AAA	CGT	GTT	TTG	ATC	AAA	GAA	GAG	GAG	TAT	GAT	2256
Val	Gln	Glu	Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	
735					740					745				750		
TCT	ATT	ATA	GTA	TTC	TAT	AAC	TCG	GTC	TTC	ATG	CAG	AGA	CTG	AAA	ACA	2304
Ser	Ile	Ile	Val	Phe	Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	
				755					760					765		
AAT	ATT	TTG	CAG	TAT	GCT	TCC	ACC	AGG	CCC	CCT	ACC	TTG	TCA	CCA	ATA	2352
Asn	Ile	Leu	Gln	Tyr	Ala	Ser	Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	
			770					775					780			
CCT	CAC	ATT	CCT	CGA	AGC	CCT	TAC	AAG	TTT	CCT	AGT	TCA	CCC	TTA	CGG	2400
Pro	His	Ile	Pro	Arg	Ser	Pro	Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	
		785					790					795				

ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys 800 805 810	2448
ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg 815 820 825 830	2496
ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln 835 840 845	2544
AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser 850 855 860	2592
GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp 865 870 875	2640
ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu 880 885 890	2688
TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg 895 900 905 910	2736
ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu 915 920 925	2784
GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG Glu Lys	2840
TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC	2900
TTCAGCTCTT TTTGTGGATA TAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC	2960
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA	3020
TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT	3080
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT	3140
TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT	3200
TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTTGA	3260
TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAAATGGAT ATTATTAGAA	3320
ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT	3380

ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT 3440
AACCATATGA TACTATCATA CTA CTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT 3500
TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGTC 3555

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 928 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	1	5	10	15
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp	20	25	30	
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	35	40	45	
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	50	55	60	
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	65	70	75	80
Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	85	90	95	
Glu	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu	100	105	110	
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	115	120	125	
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	130	135	140	
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	145	150	155	160
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	165	170	175	
Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys	180	185	190	

Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	
				485					490					495		
Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	
			500					505					510			
Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	
		515					520					525				
Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	
	530					535					540					
Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	
545					550					555					560	
Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	
				565					570					575		
Lys	Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	
			580					585					590			
Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	
		595					600					605				
Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	
	610					615					620					
Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	
625					630					635					640	
Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	
				645					650					655		
Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	
			660					665					670			
His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	
		675					680					685				
Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	
	690					695					700					
Met	Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	
705					710					715					720	
Phe	Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	
				725					730					735		
Glu	Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	
			740					745					750			
Ile	Val	Phe	Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	
		755					760						765			

Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His
770 775 780

Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro
785 790 795 800

Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser
805 810 815

Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu
820 825 830

Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile
835 840 845

Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu
850 855 860

Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu
865 870 875 880

Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys
885 890 895

Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln
900 905 910

Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
915 920 925

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3218 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2454

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCCGTC ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	48
Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
1 5 10	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	96
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
15 20 25 30	

AAA Lys	GTT Val	GAT Asp	AAT Asn	GCT Ala	ATG Met	TCA Ser	AGA Arg	CTG Leu	TTG Leu	AAG Lys	AAG Lys	TAT Tyr	GAT Asp	GTA Val	TTG Leu	144	
			35						40						45		
TTT Phe	GCA Ala	CTC Leu	TTC Phe	AGC Ser	AAA Lys	TTG Leu	GAA Glu	AGG Arg	ACA Thr	TGT Cys	GAA Glu	CTT Leu	ATA Ile	TAT Tyr	TTG Leu	192	
			50						55						60		
ACA Thr	CAA Gln	CCC Pro	AGC Ser	AGT Ser	TCG Ser	ATA Ile	TCT Ser	ACT Thr	GAA Glu	ATA Ile	AAT Asn	TCT Ser	GCA Ala	TTG Leu	GTG Val	240	
			65						70						75		
CTA Leu	AAA Lys	GTT Val	TCT Ser	TGG Trp	ATC Ile	ACA Thr	TTT Phe	TTA Leu	TTA Leu	GCT Ala	AAA Lys	GGG Gly	GAA Glu	GTA Val	TTA Leu	288	
			80						85						90		
CAA Gln	ATG Met	GAA Glu	GAT Asp	GAT Asp	CTG Leu	GTG Val	ATT Ile	TCA Ser	TTT Phe	CAG Gln	TTA Leu	ATG Met	CTA Leu	TGT Cys	GTC Val	336	
			95						100						105		
CTT Leu	GAC Asp	TAT Tyr	TTT Phe	ATT Ile	AAA Lys	CTC Leu	TCA Ser	CCT Pro	CCC Pro	ATG Met	TTG Leu	CTC Leu	AAA Lys	GAA Glu	CCA Pro	384	
			115						120						125		
TAT Tyr	AAA Lys	ACA Thr	GCT Ala	GTT Val	ATA Ile	CCC Pro	ATT Ile	AAT Asn	GGT Gly	TCA Ser	CCT Pro	CGA Arg	ACA Thr	CCC Pro	AGG Arg	432	
			130						135						140		
CGA Arg	GGT Gly	CAG Gln	AAC Asn	AGG Arg	AGT Ser	GCA Ala	CGG Arg	ATA Ile	GCA Ala	AAA Lys	CAA Gln	CTA Leu	GAA Glu	AAT Asn	GAT Asp	480	
			145						150						155		
ACA Thr	AGA Arg	ATT Ile	ATT Ile	GAA Glu	GTT Val	CTC Leu	TGT Cys	AAA Lys	GAA Glu	CAT His	GAA Glu	TGT Cys	AAT Asn	ATA Ile	GAT Asp	528	
			160						165						170		
GAG Glu	GTG Val	AAA Lys	AAT Asn	GTT Val	TAT Tyr	TTC Phe	AAA Lys	AAT Asn	TTT Phe	ATA Ile	CCT Pro	TTT Phe	ATG Met	AAT Asn	TCT Ser	576	
			175						180						185		
CTT Leu	GGA Gly	CTT Leu	GTA Val	ACA Thr	TCT Ser	AAT Asn	GGA Gly	CTT Leu	CCA Pro	GAG Glu	GTT Val	GAA Glu	AAT Asn	CTT Leu	TCT Ser	624	
			195						200						205		
AAA Lys	CGA Arg	TAC Tyr	GAA Glu	GAA Glu	ATT Ile	TAT Tyr	CTT Leu	AAA Lys	AAT Asn	AAA Lys	GAT Asp	CTA Leu	GAT Asp	GCA Ala	AGA Arg	672	
			210						215						220		
TTA Leu	TTT Phe	TTG Leu	GAT Asp	CAT His	GAT Asp	AAA Lys	ACT Thr	CTT Leu	CAG Gln	ACT Thr	GAT Asp	TCT Ser	ATA Ile	GAC Asp	AGT Ser	720	
			225						230						235		

TTT GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val 240 245 250	768
AAT GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile 255 260 265 270	816
CAA CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu 275 280 285	864
AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser 290 295 300	912
ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe 305 310 315	960
GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr 320 325 330	1008
AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys 335 340 345 350	1056
TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn 355 360 365	1104
GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val 370 375 380	1152
ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr 385 390 395	1200
GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe 400 405 410	1248
GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu 415 420 425 430	1296
ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met 435 440 445	1344

GAA Glu	TCC Ser	CTT Leu	GCA Ala	TGG Trp	CTC Leu	TCA Ser	GAT Asp	TCA Ser	CCT Pro	TTA Leu	TTT Phe	GAT Asp	CTT Leu	ATT Ile	AAA Lys	1392
			450				455				460					
CAA Gln	TCA Ser	AAG Lys	GAC Asp	CGA Arg	GAA Glu	GGA Gly	CCA Pro	ACT Thr	GAT Asp	CAC His	CTT Leu	GAA Glu	TCT Ser	GCT Ala	TGT Cys	1440
			465				470				475					
CCT Pro	CTT Leu	AAT Asn	CTT Leu	CCT Pro	CTC Leu	CAG Gln	AAT Asn	AAT Asn	CAC His	ACT Thr	GCA Ala	GCA Ala	GAT Asp	ATG Met	TAT Tyr	1488
			480				485				490					
CTT Leu	TCT Ser	CCT Pro	GTA Val	AGA Arg	TCT Ser	CCA Pro	AAG Lys	AAA Lys	AAA Lys	GGT Gly	TCA Ser	ACT Thr	ACG Thr	CGT Arg	GTA Val	1536
			495				500				505					
AAT Asn	TCT Ser	ACT Thr	GCA Ala	AAT Asn	GCA Ala	GAG Glu	ACA Thr	CAA Gln	GCA Ala	ACC Thr	TCA Ser	GCC Ala	TTC Phe	CAG Gln	ACC Thr	1584
			515				520				525					
CAG Gln	AAG Lys	CCA Pro	TTG Leu	AAA Lys	TCT Ser	ACC Thr	TCT Ser	CTT Leu	TCA Ser	CTG Leu	TTT Phe	TAT Tyr	AAA Lys	AAA Lys	GTG Val	1632
			530				535				540					
TAT Tyr	CGG Arg	CTA Leu	GCC Ala	TAT Tyr	CTC Leu	CGG Arg	CTA Leu	AAT Asn	ACA Thr	CTT Leu	TGT Cys	GAA Glu	CGC Arg	CTT Leu	CTG Leu	1680
			545				550				555					
TCT Ser	GAG Glu	CAC His	CCA Pro	GAA Glu	TTA Leu	GAA Glu	CAT His	ATC Ile	ATC Ile	TGG Trp	ACC Thr	CTT Leu	TTC Phe	CAG Gln	CAC His	1728
			560				565				570					
ACC Thr	CTG Leu	CAG Gln	AAT Asn	GAG Glu	TAT Tyr	GAA Glu	CTC Leu	ATG Met	AGA Arg	GAC Asp	AGG Arg	CAT His	TTG Leu	GAC Asp	CAA Gln	1776
			575				580				585					
ATT Ile	ATG Met	ATG Met	TGT Cys	TCC Ser	ATG Met	TAT Tyr	GGC Gly	ATA Ile	TGC Cys	AAA Lys	GTG Val	AAG Lys	AAT Asn	ATA Ile	GAC Asp	1824
			595				600				605					
CTT Leu	AAA Lys	TTC Phe	AAA Lys	ATC Ile	ATT Ile	GTA Val	ACA Thr	GCA Ala	TAC Tyr	AAG Lys	GAT Asp	CTT Leu	CCT Pro	CAT His	GCT Ala	1872
			610				615				620					
GTT Val	CAG Gln	GAG Glu	ACA Thr	TTC Phe	AAA Lys	CGT Arg	GTT Val	TTG Leu	ATC Ile	AAA Lys	GAA Glu	GAG Glu	GAG Glu	TAT Tyr	GAT Asp	1920
			625				630				635					
TCT Ser	ATT Ile	ATA Ile	GTA Val	TTC Phe	TAT Tyr	AAC Asn	TCG Ser	GTC Val	TTC Phe	ATG Met	CAG Gln	AGA Arg	CTG Leu	AAA Lys	ACA Thr	1968
			640				645				650					

AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile 655 660 665 670	2016
CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg 675 680 685	2064
ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys 690 695 700	2112
ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg 705 710 715	2160
ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln 720 725 730	2208
AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser 735 740 745 750	2256
GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp 755 760 765	2304
ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu 770 775 780	2352
TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg 785 790 795	2400
ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu 800 805 810	2448
GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG Glu Lys 815	2504
TCTCTCACAG ATGTGACTGT ATAACTTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC	2564
TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC	2624
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA	2684
TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT	2744
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT	2804

TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT 2864
TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCCTATCT ATCTTCCAAA TGCAATTTGA 2924
TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA 2984
ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT 3044
ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT 3104
AACCATATGA TACTATCATA CTACTGAAAC AGATTTTATA CCTCAGAATG TAAAAGAACT 3164
TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT 3218

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 816 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	1	5	10	15
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	20	25	30	
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	35	40	45	
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	50	55	60	
Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys	65	70	75	80
Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	Met	85	90	95	
Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp	100	105	110	
Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys	115	120	125	
Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly	130	135	140	

Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser	435	440	445
Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser	450	455	460
Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu	465	470	475
Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser	485	490	495
Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser	500	505	510
Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys	515	520	525
Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg	530	535	540
Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu	545	550	555
His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu	565	570	575
Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met	580	585	590
Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys	595	600	605
Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln	610	615	620
Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile	625	630	635
Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile	645	650	655
Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His	660	665	670
Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro	675	680	685
Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser	690	695	700
Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu	705	710	715
			720

Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile
725 730 735

Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu
740 745 750

Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu
755 760 765

Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys
770 775 780

Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln
785 790 795 800

Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
805 810 815

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 285 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTCGAGCAAT GGGCGTGATA GCGGTTTGAC TCACGGGGAT TTCCAAGTCT CCACCCCAT	60
GACGTCAATG GGAGTTTGTT TTGGCACCAA AATCAACGGG ACTTTCCAAA ATGTCGTAAC	120
AACTCCGCCC CATTGACGCA AATGGGCGGT AGGCGTGTAC GGTGGGAGGT CTATATAAGC	180
AGAGCTCGTT TAGTGAACCG TCAGATCGCC TGGAGACGCC ATCCACGCTG TTTTGACCTC	240
CATAGAAGAC ACCGGGACCG ATCCAGCCTC CGCGGCCGCG AATTC	285

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGCTCGAGC AATGGGCGTG GATAGCGG	28
--------------------------------	----

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCGCTCGAGC ACCAAAATCA ACGGGA

26

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCGCTCGAGC AACTCCGCCC CATTGAC

27

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TAGACATATG AATTCGCGGC C

21

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTAGAATTCTG CTGTCTGCG

19

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCTCTAGATG CAGTTGGACC TGGGAG

26

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCCAAGCTTG CCGCCATGTC GTTCACTTTT AC

32

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTCCAAGAGA ATTCATAAAA GG

22

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCCAAGCTTG CCGCCATGGA GCAGGACAGC GGCCCGGAC

39

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCCAAGCTTG CCGCCATGGA TTTTACTGCA TTATGTCAG

39

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCCAAGCTTG CCGCCATGGA GAAAGTTTCA TCTTGTGAT

39

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCCAAGCTTG CCGCCATGCT GTGGGGAATC TGTATCTTT

39

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCCAAGCTTG CCGCCATGTC AAGACTGTTG AAGAAG

36

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCGCCTGAGG ACCTAGATGA GATGTCGTTC

30

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCGGTTAACC CTAGATGAGA TGTCGTTTAC T

31

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCCAAGCTTG CCGTCATGCC GCCCAAAACC CCCC GA

36

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTCACCTAGG TCAACTGCTG CAAT

24

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTTGACCTAG GTGATATGTC GTTC

24

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCGCCTAGGA TCTACTGAAA TAAATTCTGC A

31

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCCGATATCA ACTGCTGGGT TGTGTCAAAT A

31

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCCGAATTCG TTTTATATGG TTCTTTGAGC AA

32

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 4..5

(D) OTHER INFORMATION: /note= "R=A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCCRCCAUGG

10

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3455 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 7..2691

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCCGTC ATG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG CTT	48
Met Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg Leu	
1 5 10	
GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG AAA	96
Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln Lys	
15 20 25 30	
TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG GAG	144
Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp Glu	
35 40 45	
AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG AAA	192
Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys Lys	
50 55 60	
AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA GAT	240
Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Asp	
65 70 75	
GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC AGT	288
Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile Ser	
80 85 90	
GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC AAA	336
Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr Lys	
95 100 105 110	
GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG TTT	384
Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu Phe	
115 120 125	

GCA Ala	CTC Leu	TTT Phe	AGC Ser	AAA Lys	TTG Leu	GAA Glu	AGG Arg	ACA Thr	TGT Cys	GAA Glu	CTT Leu	ATA Ile	TAT Tyr	TTG Leu	ACA Thr	432
			130			135						140				
CAA Gln	CCC Pro	AGC Ser	AGT Ser	TCG Ser	ATA Ile	TCT Ser	ACT Thr	GAA Glu	ATA Ile	AAT Asn	TCT Ser	GCA Ala	TTG Leu	GTG Val	CTA Leu	480
			145			150						155				
AAA Lys	GTT Val	TCT Ser	TGG Trp	ATC Ile	ACA Thr	TTT Phe	TTA Leu	TTA Leu	GCT Ala	AAA Lys	GGG Gly	GAA Glu	GTA Val	TTA Leu	CAA Gln	528
			160			165						170				
ATG Met	GAA Glu	GAT Asp	GAT Asp	CTG Leu	GTG Val	ATT Ile	TCA Ser	TTT Phe	CAG Gln	TTA Leu	ATG Met	CTA Leu	TGT Cys	GTC Val	CTT Leu	576
175						180			185						190	
GAC Asp	TAT Tyr	TTT Phe	ATT Ile	AAA Lys	CTC Leu	TCA Ser	CCT Pro	CCC Pro	ATG Met	TTG Leu	CTC Leu	AAA Lys	GAA Glu	CCA Pro	TAT Tyr	624
			195						200						205	
AAA Lys	ACA Thr	GCT Ala	GTT Val	ATA Ile	CCC Pro	ATT Ile	AAT Asn	GGT Gly	TCA Ser	CCT Pro	CGA Arg	ACA Thr	CCC Pro	AGG Arg	CGA Arg	672
			210			215						220				
GGT Gly	CAG Gln	AAC Asn	AGG Arg	AGT Ser	GCA Ala	CGG Arg	ATA Ile	GCA Ala	AAA Lys	CAA Gln	CTA Leu	GAA Glu	AAT Asn	GAT Asp	ACA Thr	720
			225			230						235				
AGA Arg	ATT Ile	ATT Ile	GAA Glu	GTT Val	CTC Leu	TGT Cys	AAA Lys	GAA Glu	CAT His	GAA Glu	TGT Cys	AAT Asn	ATA Ile	GAT Asp	GAG Glu	768
240						245						250				
GTG Val	AAA Lys	AAT Asn	GTT Val	TAT Tyr	TTC Phe	AAA Lys	AAT Asn	TTT Phe	ATA Ile	CCT Pro	TTT Phe	ATG Met	AAT Asn	TCT Ser	CTT Leu	816
255						260			265						270	
GGA Gly	CTT Leu	GTA Val	ACA Thr	TCT Ser	AAT Asn	GGA Gly	CTT Leu	CCA Pro	GAG Glu	GTT Val	GAA Glu	AAT Asn	CTT Leu	TCT Ser	AAA Lys	864
			275						280						285	
CGA Arg	TAC Tyr	GAA Glu	GAA Glu	ATT Ile	TAT Tyr	CTT Leu	AAA Lys	AAT Asn	AAA Lys	GAT Asp	CTA Leu	GAT Asp	GCA Ala	AGA Arg	TTA Leu	912
			290			295						300				
TTT Phe	TTG Leu	GAT Asp	CAT His	GAT Asp	AAA Lys	ACT Thr	CTT Leu	CAG Gln	ACT Thr	GAT Asp	TCT Ser	ATA Ile	GAC Asp	AGT Ser	TTT Phe	960
305						310						315				
GAA Glu	ACA Thr	CAG Gln	AGA Arg	ACA Thr	CCA Pro	CGA Arg	AAA Lys	AGT Ser	AAC Asn	CTT Leu	GAT Asp	GAA Glu	GAG Glu	GTG Val	AAT Asn	1008
320						325			330							

GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	AAC	ACT	ATC	CAA	1056
Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	
335					340					345					350	
CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	CCT	TCA	GAA	AAT	1104
Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	
				355					360					365		
CTG	ATT	TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	AGT	ATA	1152
Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	
			370					375					380			
CTG	AAA	AGA	GTG	AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	GAG	AAA	TTT	GCT	1200
Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	
		385					390					395				
AAA	GCT	GTG	GGA	CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	CAG	CGA	TAC	AAA	1248
Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	
	400					405					410					
CTT	GGA	GTT	CGC	TTG	TAT	TAC	CGA	GTA	ATG	GAA	TCC	ATG	CTT	AAA	TCA	1296
Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	
415					420					425					430	
GAA	GAA	GAA	CGA	TTA	TCC	ATT	CAA	AAT	TTT	AGC	AAA	CTT	CTG	AAT	GAC	1344
Glu	Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	
			435					440					445			
AAC	ATT	TTT	CAT	ATG	TCT	TTA	TTG	GCG	TGC	GCT	CTT	GAG	GTT	GTA	ATG	1392
Asn	Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	
			450					455				460				
GCC	ACA	TAT	AGC	AGA	AGT	ACA	TCT	CAG	AAT	CTT	GAT	TCT	GGA	ACA	GAT	1440
Ala	Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	
		465					470					475				
TTG	TCT	TTC	CCA	TGG	ATT	CTG	AAT	GTG	CTT	AAT	TTA	AAA	GCC	TTT	GAT	1488
Leu	Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	
	480					485					490					
TTT	TAC	AAA	GTG	ATC	GAA	AGT	TTT	ATC	AAA	GCA	GAA	GGC	AAC	TTG	ACA	1536
Phe	Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	
495					500					505					510	
AGA	GAA	ATG	ATA	AAA	CAT	TTA	GAA	CGA	TGT	GAA	CAT	CGA	ATC	ATG	GAA	1584
Arg	Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	
				515				520					525			
TCC	CTT	GCA	TGG	CTC	TCA	GAT	TCA	CCT	TTA	TTT	GAT	CTT	ATT	AAA	CAA	1632
Ser	Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	
			530					535					540			

TCA Ser	AAG Lys	GAC Asp	CGA Arg	GAA Glu	GGA Gly	CCA Pro	ACT Thr	GAT Asp	CAC His	CTT Leu	GAA Glu	TCT Ser	GCT Ala	TGT Cys	CCT Pro	1680
545						550						555				
CTT Leu	AAT Asn	CTT Leu	CCT Pro	CTC Leu	CAG Gln	AAT Asn	AAT Asn	CAC His	ACT Thr	GCA Ala	GCA Ala	GAT Asp	ATG Met	TAT Tyr	CTT Leu	1728
560						565						570				
TCT Ser	CCT Pro	GTA Val	AGA Arg	TCT Ser	CCA Pro	AAG Lys	AAA Lys	AAA Lys	GGT Gly	TCA Ser	ACT Thr	ACG Thr	CGT Arg	GTA Val	AAT Asn	1776
575						580						585			590	
TCT Ser	ACT Thr	GCA Ala	AAT Asn	GCA Ala	GAG Glu	ACA Thr	CAA Gln	GCA Ala	ACC Thr	TCA Ser	GCC Ala	TTC Phe	CAG Gln	ACC Thr	CAG Gln	1824
			595						600						605	
AAG Lys	CCA Pro	TTG Leu	AAA Lys	TCT Ser	ACC Thr	TCT Ser	CTT Leu	TCA Ser	CTG Leu	TTT Phe	TAT Tyr	AAA Lys	AAA Lys	GTG Val	TAT Tyr	1872
			610						615						620	
CGG Arg	CTA Leu	GCC Ala	TAT Tyr	CTC Leu	CGG Arg	CTA Leu	AAT Asn	ACA Thr	CTT Leu	TGT Cys	GAA Glu	CGC Arg	CTT Leu	CTG Leu	TCT Ser	1920
			625						630						635	
GAG Glu	CAC His	CCA Pro	GAA Glu	TTA Leu	GAA Glu	CAT His	ATC Ile	ATC Ile	TGG Trp	ACC Thr	CTT Leu	TTC Phe	CAG Gln	CAC His	ACC Thr	1968
			640			645						650				
CTG Leu	CAG Gln	AAT Asn	GAG Glu	TAT Tyr	GAA Glu	CTC Leu	ATG Met	AGA Arg	GAC Asp	AGG Arg	CAT His	TTG Leu	GAC Asp	CAA Gln	ATT Ile	2016
655						660						665			670	
ATG Met	ATG Met	TGT Cys	TCC Ser	ATG Met	TAT Tyr	GGC Gly	ATA Ile	TGC Cys	AAA Lys	GTG Val	AAG Lys	AAT Asn	ATA Ile	GAC Asp	CTT Leu	2064
			675						680						685	
AAA Lys	TTC Phe	AAA Lys	ATC Ile	ATT Ile	GTA Val	ACA Thr	GCA Ala	TAC Tyr	AAG Lys	GAT Asp	CTT Leu	CCT Pro	CAT His	GCT Ala	GTT Val	2112
			690						695						700	
CAG Gln	GAG Glu	ACA Thr	TTC Phe	AAA Lys	CGT Arg	GTT Val	TTG Leu	ATC Ile	AAA Lys	GAA Glu	GAG Glu	GAG Glu	TAT Tyr	GAT Asp	TCT Ser	2160
			705			710						715				
ATT Ile	ATA Ile	GTA Val	TTC Phe	TAT Tyr	AAC Asn	TCG Ser	GTC Val	TTC Phe	ATG Met	CAG Gln	AGA Arg	CTG Leu	AAA Lys	ACA Thr	AAT Asn	2208
720						725						730				
ATT Ile	TTG Leu	CAG Gln	TAT Tyr	GCT Ala	TCC Ser	ACC Thr	AGG Arg	CCC Pro	CCT Pro	ACC Thr	TTG Leu	TCA Ser	CCA Pro	ATA Ile	CCT Pro	2256
735						740						745			750	

CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG ATT	2304
His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile	
755 760 765	
CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA ATT	2352
Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile	
770 775 780	
TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA ATC	2400
Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile	
785 790 795	
TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG AAA	2448
Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys	
800 805 810	
ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT GCT	2496
Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala	
815 820 825 830	
GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT ATT	2544
Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile	
835 840 845	
GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG TCC	2592
Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser	
850 855 860	
AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA ATG	2640
Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met	
865 870 875	
CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG	2688
Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu	
880 885 890	
AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG	2741
Lys	
895	
TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC	2801
TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC	2861
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA	2921
TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT	2981
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT	3041
TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT	3101
TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCCTATCT ATCTTCCAAA TGCAATTGA	3161

TTGACTGCCC ATTACACAAA ATTATCCTGA ACTCTTCTGC AAAAAATGGAT ATTATTAGAA 3221
ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT 3281
ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT 3341
AACCATATGA TACTATCATA CTA CTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT 3401
TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT 3455

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 895 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg Leu Glu Phe
1 5 10 15
Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln Lys Leu Lys
20 25 30
Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp Glu Lys Val
35 40 45
Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys Lys Lys Glu
50 55 60
Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Asp Glu Met
65 70 75 80
Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile Ser Val His
85 90 95
Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr Lys Val Asp
100 105 110
Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala Leu
115 120 125
Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln Pro
130 135 140
Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val
145 150 155 160
Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met Glu
165 170 175

Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp	Tyr	180	185	190
Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys	Thr	195	200	205
Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly	Gln	210	215	220
Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	Arg	Ile	225	230	235
Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu	Val	Lys	245	250	255
Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu	Gly	Leu	260	265	270
Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	Lys	Arg	Tyr	275	280	285
Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	Leu	Phe	Leu	290	295	300
Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	Phe	Glu	Thr	305	310	315
Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	Asn	Val	Ile	325	330	335
Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln	Leu	340	345	350
Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu	Ile	355	360	365
Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu	Lys	370	375	380
Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys	Ala	385	390	395
Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	Gly	405	410	415
Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu	Glu	420	425	430
Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	Asn	Ile	435	440	445
Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	Thr	450	455	460

Tyr 465	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	Ser	480		
					470								475					480
Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	Tyr			
					485								490					495
Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	Glu			
					500								505					510
Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	Leu			
					515								520					525
Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	Lys			
					530								535					540
Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	Asn			
545					550								555					560
Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	Pro			
					565								570					575
Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	Thr			
					580								585					590
Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	Pro			
					595								600					605
Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	Leu			
					610								615					620
Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	His			
625					630								635					640
Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	Gln			
					645								650					655
Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	Met			
					660								665					670
Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	Phe			
					675								680					685
Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	Glu			
					690								695					700
Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	Ile			
705					710								715					720
Val	Phe	Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	Leu			
					725								730					735
Gln	Tyr	Ala	Ser	Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	Ile			
					740								745					750

Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly
755 760 765

Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu
770 775 780

Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val
785 790 795 800

Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn
805 810 815

Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly
820 825 830

Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly
835 840 845

Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe
850 855 860

Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys
865 870 875 880

Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
885 890 895

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3392 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2628

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GCCATC ATG GAT TTT ACT GCA TTA TGT CAG AAA TTA AAG ATA CCA GAT	48
Met Asp Phe Thr Ala Leu Cys Gln Lys Leu Lys Ile Pro Asp	
1 5 10	
CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG GAG AAA GTT TCA TCT GTG	96
His Val Arg Glu Arg Ala Trp Leu Thr Trp Glu Lys Val Ser Ser Val	
15 20 25 30	
GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG AAA AAG GAA CTG TGG GGA	144
Asp Gly Val Leu Gly Tyr Ile Gln Lys Lys Lys Glu Leu Trp Gly	
35 40 45	

ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA GAT GAG ATG TCG TTC ACT	192
Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Asp Glu Met Ser Phe Thr	
50 55 60	
TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC AGT GTC CAT AAA TTC TTT	240
Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile Ser Val His Lys Phe Phe	
65 70 75	
AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC AAA GTT GAT AAT GCT ATG	288
Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr Lys Val Asp Asn Ala Met	
80 85 90	
TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG TTT GCA CTC TTC AGC AAA	336
Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala Leu Phe Ser Lys	
95 100 105 110	
TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG ACA CAA CCC AGC AGT TCG	384
Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln Pro Ser Ser Ser	
115 120 125	
ATA TCT ACT GAA ATA AAT TCT GCA TTG GTG CTA AAA GTT TCT TGG ATC	432
Ile Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val Ser Trp Ile	
130 135 140	
ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA CAA ATG GAA GAT GAT CTG	480
Thr Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met Glu Asp Asp Leu	
145 150 155	
GTG ATT TCA TTT CAG TTA ATG CTA TGT GTC CTT GAC TAT TTT ATT AAA	528
Val Ile Ser Phe Gln Leu Met Leu Cys Val Leu Asp Tyr Phe Ile Lys	
160 165 170	
CTC TCA CCT CCC ATG TTG CTC AAA GAA CCA TAT AAA ACA GCT GTT ATA	576
Leu Ser Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys Thr Ala Val Ile	
175 180 185 190	
CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG CGA GGT CAG AAC AGG AGT	624
Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg Ser	
195 200 205	
GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT ACA AGA ATT ATT GAA GTT	672
Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu Val	
210 215 220	
CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT GAG GTG AAA AAT GTT TAT	720
Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val Tyr	
225 230 235	
TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT CTT GGA CTT GTA ACA TCT	768
Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr Ser	
240 245 250	

AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT AAA CGA TAC GAA GAA ATT	816
Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile	
255 260 265 270	
TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA TTA TTT TTG GAT CAT GAT	864
Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp	
275 280 285	
AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT TTT GAA ACA CAG AGA ACA	912
Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr	
290 295 300	
CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG AAT GTA ATT CCT CCA CAC	960
Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro His	
305 310 315	
ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC CAA CAA TTA ATG ATG ATT	1008
Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met Ile	
320 325 330	
TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA AAT CTG ATT TCC TAT TTT	1056
Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe	
335 340 345 350	
AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT ATA CTG AAA AGA GTG AAG	1104
Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys	
355 360 365	
GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT GCT AAA GCT GTG GGA CAG	1152
Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln	
370 375 380	
GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC AAA CTT GGA GTT CGC TTG	1200
Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu	
385 390 395	
TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA TCA GAA GAA GAA CGA TTA	1248
Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu	
400 405 410	
TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT GAC AAC ATT TTT CAT ATG	1296
Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met	
415 420 425 430	
TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA ATG GCC ACA TAT AGC AGA	1344
Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg	
435 440 445	
AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA GAT TTG TCT TTC CCA TGG	1392
Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp	
450 455 460	

ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT GAT TTT TAC AAA GTG ATC Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile 465 470 475	1440
GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG ACA AGA GAA ATG ATA AAA Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys 480 485 490	1488
CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG GAA TCC CTT GCA TGG CTC His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu 495 500 505 510	1536
TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA CAA TCA AAG GAC CGA GAA Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu 515 520 525	1584
GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT CCT CTT AAT CTT CCT CTC Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu 530 535 540	1632
CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT CTT TCT CCT GTA AGA TCT Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser 545 550 555	1680
CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT GCA AAT GCA Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala 560 565 570	1728
GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA TTG AAA TCT Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys Ser 575 580 585 590	1776
ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA GCC TAT CTC Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr Leu 595 600 605	1824
CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG TCT GAG CAC CCA GAA TTA Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His Pro Glu Leu 610 615 620	1872
GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC ACC CTG CAG AAT GAG TAT Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln Asn Glu Tyr 625 630 635	1920
GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA ATT ATG ATG TGT TCC ATG Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met Met Cys Ser Met 640 645 650	1968
TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC CTT AAA TTC AAA ATC ATT Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys Phe Lys Ile Ile 655 660 665 670	2016

GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT GTT CAG GAG ACA TTC AAA	2064
Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln Glu Thr Phe Lys	
675 680 685	
CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT TCT ATT ATA GTA TTC TAT	2112
Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe Tyr	
690 695 700	
AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA AAT ATT TTG CAG TAT GCT	2160
Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr Ala	
705 710 715	
TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA CCT CAC ATT CCT CGA AGC	2208
Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His Ile Pro Arg Ser	
720 725 730	
CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG ATT CCT GGA GGG AAC ATC	2256
Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn Ile	
735 740 745 750	
TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA ATT TCA GAA GGT CTG CCA	2304
Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu Pro	
755 760 765	
ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA ATC TTA GTA TCA ATT GGT	2352
Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val Ser Ile Gly	
770 775 780	
GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG AAA ATA AAT CAG ATG GTA	2400
Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn Gln Met Val	
785 790 795	
TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT GCT GAA GGA AGC AAC CCT	2448
Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly Ser Asn Pro	
800 805 810	
CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT ATT GAA GGA TCA GAT GAA	2496
Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp Glu	
815 820 825 830	
GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG TCC AAA TTT CAG CAG AAA	2544
Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln Lys	
835 840 845	
CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA ATG CAA AAG CAG AAA ATG	2592
Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys Met	
850 855 860	
AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG AAA TGAGGATCTC	2638
Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys	
865 870	
AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG TCTCTCACAG ATGTGACTGT	2698

ATAACTTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT TTTGTGGATA 2758
TAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG TTAGTCATTG 2818
TTATTTATAC AAGATTGAAA ATCTTG TGTA AATCCTGCCA TTTAAAAAGT TG TAGCAGAT 2878
TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG GCCCTAGAGT 2938
GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG CATATAGGTG 2998
ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC ATGAACACCC 3058
TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTGA TTGACTGCCC ATTCACCAAA 3118
ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA AATTACTAAT 3178
TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT ACTGTGTGCT TGTTTATAA 3238
AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA TACTATCATA 3298
CTACTGAAAC AGATTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT TTTCTTCATC 3358
CAACTTATGT TTTTAAATGA GGATTATTGA TAGT 3392

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 874 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	Lys	Ile	Pro	Asp	His	Val
1				5					10					15	
Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	Val	Ser	Ser	Val	Asp	Gly
			20					25					30		
Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	Glu	Leu	Trp	Gly	Ile	Cys
		35					40					45			
Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu	Met	Ser	Phe	Thr	Phe	Thr
	50					55					60				
Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	His	Lys	Phe	Phe	Asn	Leu
65					70					75				80	
Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	Asp	Asn	Ala	Met	Ser	Arg
				85					90					95	

Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala Leu Phe Ser Lys Leu Glu
 100 105 110
 Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln Pro Ser Ser Ser Ile Ser
 115 120 125
 Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val Ser Trp Ile Thr Phe
 130 135 140
 Leu Leu Ala Lys Gly Glu Val Leu Gln Met Glu Asp Asp Leu Val Ile
 145 150 155 160
 Ser Phe Gln Leu Met Leu Cys Val Leu Asp Tyr Phe Ile Lys Leu Ser
 165 170 175
 Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys Thr Ala Val Ile Pro Ile
 180 185 190
 Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg Ser Ala Arg
 195 200 205
 Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu Val Leu Cys
 210 215 220
 Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val Tyr Phe Lys
 225 230 235 240
 Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr Ser Asn Gly
 245 250 255
 Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile Tyr Leu
 260 265 270
 Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp Lys Thr
 275 280 285
 Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro Arg
 290 295 300
 Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr Pro
 305 310 315 320
 Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met Ile Leu Asn
 325 330 335
 Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn
 340 345 350
 Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile
 355 360 365
 Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys
 370 375 380

Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	Gly	Val	Arg	Leu	Tyr	Tyr		
385					390					395					400		
Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu	Glu	Glu	Arg	Leu	Ser	Ile		
				405					410					415			
Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	Asn	Ile	Phe	His	Met	Ser	Leu		
			420					425					430				
Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	Thr	Tyr	Ser	Arg	Ser	Thr		
	435						440					445					
Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	Ser	Phe	Pro	Trp	Ile	Leu		
450						455					460						
Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	Tyr	Lys	Val	Ile	Glu	Ser		
465					470					475					480		
Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	Glu	Met	Ile	Lys	His	Leu		
				485					490					495			
Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	Leu	Ala	Trp	Leu	Ser	Asp		
			500					505					510				
Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	Lys	Asp	Arg	Glu	Gly	Pro		
	515						520					525					
Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	Asn	Leu	Pro	Leu	Gln	Asn		
	530					535					540						
Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	Pro	Val	Arg	Ser	Pro	Lys		
545					550					555					560		
Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	Thr	Ala	Asn	Ala	Glu	Thr		
			565						570					575			
Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	Pro	Leu	Lys	Ser	Thr	Ser		
			580					585					590				
Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	Leu	Ala	Tyr	Leu	Arg	Leu		
	595					600					605						
Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	His	Pro	Glu	Leu	Glu	His		
	610					615					620						
Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	Gln	Asn	Glu	Tyr	Glu	Leu		
625					630					635					640		
Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	Met	Cys	Ser	Met	Tyr	Gly		
				645					650					655			
Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	Phe	Lys	Ile	Ile	Val	Thr		
			660					665					670				

Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	Glu	Thr	Phe	Lys	Arg	Val	
	675						680					685				
Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	Ile	Val	Phe	Tyr	Asn	Ser	
	690					695					700					
Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	Leu	Gln	Tyr	Ala	Ser	Thr	
	705				710					715					720	
Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	Ile	Pro	Arg	Ser	Pro	Tyr	
				725					730					735		
Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	Gly	Gly	Asn	Ile	Tyr	Ile	
			740					745					750			
Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	Glu	Gly	Leu	Pro	Thr	Pro	
		755					760					765				
Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu	Val	Ser	Ile	Gly	Glu	Ser	
	770					775					780					
Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile	Asn	Gln	Met	Val	Cys	Asn	
	785				790					795					800	
Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu	Gly	Ser	Asn	Pro	Pro	Lys	
				805					810					815		
Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu	Gly	Ser	Asp	Glu	Ala	Asp	
			820					825						830		
Gly	Ser	Lys	His	Leu	Pro	Gly	Glu	Ser	Lys	Phe	Gln	Gln	Lys	Leu	Ala	
		835					840					845				
Glu	Met	Thr	Ser	Thr	Arg	Thr	Arg	Met	Gln	Lys	Gln	Lys	Met	Asn	Asp	
	850					855					860					
Ser	Met	Asp	Thr	Ser	Asn	Lys	Glu	Glu	Lys							
	865				870											

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3323 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2559

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCCATC ATG GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT	48
Met Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr	
1 5 10	
ATT CAA AAG AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA	96
Ile Gln Lys Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala	
15 20 25 30	
GTT GAC CTA GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC	144
Val Asp Leu Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn	
35 40 45	
ATA GAA ATC AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT	192
Ile Glu Ile Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp	
50 55 60	
ACC AGT ACC AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT	240
Thr Ser Thr Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr	
65 70 75	
GAT GTA TTG TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT	288
Asp Val Leu Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu	
80 85 90	
ATA TAT TTG ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT	336
Ile Tyr Leu Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser	
95 100 105 110	
GCA TTG GTG CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG	384
Ala Leu Val Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly	
115 120 125	
GAA GTA TTA CAA ATG GAA GAT GAT CTG GTG ATT TCA TTT CAG TTA ATG	432
Glu Val Leu Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met	
130 135 140	
CTA TGT GTC CTT GAC TAT TTT ATT AAA CTC TCA CCT CCC ATG TTG CTC	480
Leu Cys Val Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu	
145 150 155	
AAA GAA CCA TAT AAA ACA GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA	528
Lys Glu Pro Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg	
160 165 170	
ACA CCC AGG CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA	576
Thr Pro Arg Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu	
175 180 185 190	
GAA AAT GAT ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT	624
Glu Asn Asp Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys	
195 200 205	

AAT Asn	ATA Ile	GAT Asp	GAG Glu 210	GTG Val	AAA Lys	AAT Asn	GTT Val	TAT Tyr 215	TTC Phe	AAA Lys	AAT Asn	TTT Phe 220	ATA Ile 220	CCT Pro	TTT Phe	672
ATG Met	AAT Asn 225	TCT Ser	CTT Leu	GGA Gly	CTT Leu	GTA Val	ACA Thr 230	TCT Ser	AAT Asn	GGA Gly	CTT Leu 235	CCA Pro 235	GAG Glu	GTT Val	GAA Glu	720
AAT Asn 240	CTT Leu	TCT Ser	AAA Lys	CGA Arg	TAC Tyr	GAA Glu 245	GAA Glu	ATT Ile	TAT Tyr	CTT Leu 250	AAA Lys 250	AAT Asn	AAA Lys	GAT Asp	CTA Leu	768
GAT Asp 255	GCA Ala	AGA Arg	TTA Leu	TTT Phe 260	TTG Leu	GAT Asp	CAT His	GAT Asp	AAA Lys 265	ACT Thr 265	CTT Leu 265	CAG Gln	ACT Thr	GAT Asp 270	TCT Ser 270	816
ATA Ile	GAC Asp	AGT Ser	TTT Phe 275	GAA Glu 275	ACA Thr	CAG Gln	AGA Arg	ACA Thr 280	CCA Pro 280	CGA Arg	AAA Lys	AGT Ser	AAC Asn	CTT Leu 285	GAT Asp	864
GAA Glu	GAG Glu	GTG Val 290	AAT Asn	GTA Val	ATT Ile	CCT Pro	CCA Pro	CAC His 295	ACT Thr	CCA Pro	GTT Val	AGG Arg 300	ACT Thr	GTT Val	ATG Met	912
AAC Asn	ACT Thr 305	ATC Ile	CAA Gln	CAA Gln	TTA Leu	ATG Met 310	ATG Met	ATT Ile	TTA Leu	AAT Asn	TCA Ser 315	GCA Ala 315	AGT Ser	GAT Asp	CAA Gln	960
CCT Pro 320	TCA Ser	GAA Glu	AAT Asn	CTG Leu	ATT Ile	TCC Ser 325	TAT Tyr	TTT Phe	AAC Asn	AAC Asn	TGC Cys 330	ACA Thr	GTG Val	AAT Asn	CCA Pro	1008
AAA Lys 335	GAA Glu	AGT Ser	ATA Ile	CTG Leu 340	AAA Lys	AGA Arg	GTG Val	AAG Lys	GAT Asp 345	ATA Ile 345	GGA Gly	TAC Tyr	ATC Ile	TTT Phe	AAA Lys 350	1056
GAG Glu	AAA Lys	TTT Phe	GCT Ala 355	AAA Lys 355	GCT Ala	GTG Val	GGA Gly	CAG Gln 360	GGT Gly 360	TGT Cys	GTC Val	GAA Glu	ATT Ile 365	GGA Gly 365	TCA Ser	1104
CAG Gln	CGA Arg	TAC Tyr 370	AAA Lys 370	CTT Leu	GGA Gly	GTT Val	CGC Arg	TTG Leu 375	TAT Tyr	TAC Tyr	CGA Arg	GTA Val 380	ATG Met	GAA Glu	TCC Ser	1152
ATG Met	CTT Leu 385	AAA Lys	TCA Ser	GAA Glu	GAA Glu	GAA Glu 390	CGA Arg	TTA Leu 390	TCC Ser	ATT Ile	CAA Gln 395	AAT Asn 395	TTT Phe	AGC Ser	AAA Lys	1200
CTT Leu 400	CTG Leu	AAT Asn	GAC Asp	AAC Asn	ATT Ile	TTT Phe 405	CAT His	ATG Met	TCT Ser	TTA Leu 410	TTG Leu 410	GCG Ala	TGC Cys	GCT Ala	CTT Leu	1248

TTG GAC CAA ATT ATG ATG TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG	1920
Leu Asp Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys	
625 630 635	
AAT ATA GAC CTT AAA TTC AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT	1968
Asn Ile Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu	
640 645 650	
CCT CAT GCT GTT CAG GAG ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG	2016
Pro His Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu	
655 660 665 670	
GAG TAT GAT TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA	2064
Glu Tyr Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg	
675 680 685	
CTG AAA ACA AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG	2112
Leu Lys Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu	
690 695 700	
TCA CCA ATA CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA	2160
Ser Pro Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser	
705 710 715	
CCC TTA CGG ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT	2208
Pro Leu Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser	
720 725 730	
CCA TAT AAA ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA	2256
Pro Tyr Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro	
735 740 745 750	
AGA TCA AGA ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG	2304
Arg Ser Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu	
755 760 765	
AAG TTC CAG AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC	2352
Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu	
770 775 780	
AAA AGA AGT GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA	2400
Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu	
785 790 795	
CGC TTT GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC	2448
Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu	
800 805 810	
CCA GGA GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT	2496
Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr	
815 820 825 830	

CGA ACA CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA	2544
Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser	
835 840 845	
AAC AAG GAA GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT	2599
Asn Lys Glu Glu Lys	
850	
GGATTCATTG TCTCTCACAG ATGTGACTGT ATAACTTTCC CAGGTTCTGT TTATGGCCAC	2659
ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA	2719
TTCCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA	2779
AATCCTGCCA TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG	2839
CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC	2899
TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT	2959
GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAATG TGCCTATCT ATCTTCCAAA	3019
TGCAATTTGA TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT	3079
ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG	3139
AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA	3199
AGCAAAGTAT AACCATATGA TACTATCATA CTAAGTAAAC AGATTTTCATA CCTCAGAATG	3259
TAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA	3319
TAGT	3323

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 851 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln	
1 5 10 15	
Lys Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp	
20 25 30	
Leu Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu	
35 40 45	

Ile Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser
 50 55 60

Thr Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val
 65 70 75 80

Leu Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr
 85 90 95

Leu Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu
 100 105 110

Val Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val
 115 120 125

Leu Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys
 130 135 140

Val Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu
 145 150 155 160

Pro Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro
 165 170 175

Arg Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn
 180 185 190

Asp Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile
 195 200 205

Asp Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn
 210 215 220

Ser Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu
 225 230 235 240

Ser Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala
 245 250 255

Arg Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp
 260 265 270

Ser Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu
 275 280 285

Val Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr
 290 295 300

Ile Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser
 305 310 315 320

Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu
 325 330 335

Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys
 340 345 350
 Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg
 355 360 365
 Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu
 370 375 380
 Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu
 385 390 395 400
 Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val
 405 410 415
 Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly
 420 425 430
 Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala
 435 440 445
 Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn
 450 455 460
 Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile
 465 470 475 480
 Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile
 485 490 495
 Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala
 500 505 510
 Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met
 515 520 525
 Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg
 530 535 540
 Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln
 545 550 555 560
 Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys
 565 570 575
 Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu
 580 585 590
 Leu Ser Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln
 595 600 605
 His Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp
 610 615 620

Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile
 625 630 635 640
 Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His
 645 650 655
 Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr
 660 665 670
 Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys
 675 680 685
 Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro
 690 695 700
 Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu
 705 710 715 720
 Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr
 725 730 735
 Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser
 740 745 750
 Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe
 755 760 765
 Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg
 770 775 780
 Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe
 785 790 795 800
 Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly
 805 810 815
 Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr
 820 825 830
 Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys
 835 840 845
 Glu Glu Lys
 850

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3266 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 7..2502

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GCCATC ATG CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	48
Met Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
1 5 10	
GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	96
Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
15 20 25 30	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	144
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
35 40 45	
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG	192
Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu	
50 55 60	
TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG	240
Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu	
65 70 75	
ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT GCA TTG GTG	288
Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val	
80 85 90	
CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA	336
Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu	
95 100 105 110	
CAA ATG GAA GAT GAT CTG GTG ATT TCA TTT CAG TTA ATG CTA TGT GTC	384
Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val	
115 120 125	
CTT GAC TAT TTT ATT AAA CTC TCA CCT CCC ATG TTG CTC AAA GAA CCA	432
Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro	
130 135 140	
TAT AAA ACA GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG	480
Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg	
145 150 155	
CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT	528
Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp	
160 165 170	
ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT	576
Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp	
175 180 185 190	

GAG	GTG	AAA	AAT	GTT	TAT	TTC	AAA	AAT	TTT	ATA	CCT	TTT	ATG	AAT	TCT	624
Glu	Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	
			195						200					205		
CTT	GGA	CTT	GTA	ACA	TCT	AAT	GGA	CTT	CCA	GAG	GTT	GAA	AAT	CTT	TCT	672
Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	
			210					215					220			
AAA	CGA	TAC	GAA	GAA	ATT	TAT	CTT	AAA	AAT	AAA	GAT	CTA	GAT	GCA	AGA	720
Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	
		225					230					235				
TTA	TTT	TTG	GAT	CAT	GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	ATA	GAC	AGT	768
Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	
	240					245					250					
TTT	GAA	ACA	CAG	AGA	ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	GAA	GAG	GTG	816
Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	
255				260					265						270	
AAT	GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	AAC	ACT	ATC	864
Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	
			275						280					285		
CAA	CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	CCT	TCA	GAA	912
Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	
			290					295					300			
AAT	CTG	ATT	TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	AGT	960
Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	
		305					310					315				
ATA	CTG	AAA	AGA	GTG	AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	GAG	AAA	TTT	1008
Ile	Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	
	320					325					330					
GCT	AAA	GCT	GTG	GGA	CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	CAG	CGA	TAC	1056
Ala	Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	
335					340				345						350	
AAA	CTT	GGA	GTT	CGC	TTG	TAT	TAC	CGA	GTA	ATG	GAA	TCC	ATG	CTT	AAA	1104
Lys	Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	
			355						360					365		
TCA	GAA	GAA	GAA	CGA	TTA	TCC	ATT	CAA	AAT	TTT	AGC	AAA	CTT	CTG	AAT	1152
Ser	Glu	Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	
			370					375					380			
GAC	AAC	ATT	TTT	CAT	ATG	TCT	TTA	TTG	GCG	TGC	GCT	CTT	GAG	GTT	GTA	1200
Asp	Asn	Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	
		385					390					395				

ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA	1248
Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr	
400 405 410	
GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT	1296
Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe	
415 420 425 430	
GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG	1344
Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu	
435 440 445	
ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG	1392
Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met	
450 455 460	
GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA	1440
Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys	
465 470 475	
CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT	1488
Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys	
480 485 490	
CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT	1536
Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr	
495 500 505 510	
CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA	1584
Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val	
515 520 525	
AAT TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC	1632
Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr	
530 535 540	
CAG AAG CCA TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG	1680
Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val	
545 550 555	
TAT CGG CTA GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG	1728
Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu	
560 565 570	
TCT GAG CAC CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC	1776
Ser Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His	
575 580 585 590	
ACC CTG CAG AAT GAG TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA	1824
Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln	
595 600 605	

ATT ATG ATG TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC	1872
Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp	
610 615 620	
CTT AAA TTC AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT	1920
Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala	
625 630 635	
GTT CAG GAG ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT	1968
Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp	
640 645 650	
TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA	2016
Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr	
655 660 665 670	
AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA	2064
Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile	
675 680 685	
CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG	2112
Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg	
690 695 700	
ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA	2160
Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys	
705 710 715	
ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA	2208
Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg	
720 725 730	
ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG	2256
Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln	
735 740 745 750	
AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT	2304
Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser	
755 760 765	
GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT	2352
Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp	
770 775 780	
ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG	2400
Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu	
785 790 795	
TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA	2448
Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg	
800 805 810	

ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA 2496
Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu
815 820 825 830

GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG 2552
Glu Lys

TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC 2612

TTCAGCTCTT TTTGTGGATA TAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC 2672

ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGCTGT AATCCTGCCA 2732

TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT 2792

AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT 2852

TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT 2912

TTAATTTAAC ATGAACACCC TTAGAAAATG TGCCTATCT ATCTTCCAAA TGCAATTTGA 2972

TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA 3032

ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT 3092

ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT 3152

AACCATATGA TACTATCATA CTACTGAAAC AGATTTTATA CCTCAGAATG TAAAGAACT 3212

TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT 3266

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 832 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Asp Glu
1 5 10 15

Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile Ser Val
20 25 30

His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr Lys Val
35 40 45

Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala		
50						55					60						
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln		
65					70					75					80		
Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys		
				85					90					95			
Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	Met		
			100					105					110				
Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp		
	115						120					125					
Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys		
	130					135					140						
Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly		
145					150					155					160		
Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	Arg		
			165						170					175			
Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu	Val		
			180					185					190				
Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu	Gly		
	195						200					205					
Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	Lys	Arg		
	210					215					220						
Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	Leu	Phe		
225					230					235					240		
Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	Phe	Glu		
				245					250					255			
Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	Asn	Val		
			260					265					270				
Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln		
	275						280					285					
Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu		
	290					295					300						
Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu		
305					310					315					320		
Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys		
				325					330					335			

Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu
 340 345 350
 Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu
 355 360 365
 Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn
 370 375 380
 Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala
 385 390 395 400
 Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu
 405 410 415
 Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe
 420 425 430
 Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg
 435 440 445
 Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser
 450 455 460
 Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser
 465 470 475 480
 Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu
 485 490 495
 Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser
 500 505 510
 Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser
 515 520 525
 Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys
 530 535 540
 Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg
 545 550 555 560
 Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu
 565 570 575
 His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu
 580 585 590
 Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met
 595 600 605
 Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys
 610 615 620

Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln
625 630 635 640
Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile
645 650 655
Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile
660 665 670
Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His
675 680 685
Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro
690 695 700
Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser
705 710 715 720
Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu
725 730 735
Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile
740 745 750
Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu
755 760 765
Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu
770 775 780
Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys
785 790 795 800
Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln
805 810 815
Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
820 825 830

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3113 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2349

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GCCGTC	ATG	TCA	AGA	CTG	TTG	AAG	AAG	TAT	GAT	GTA	TTG	TTT	GCA	CTC	48
	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	Leu	
	1				5					10					
TTC	AGC	AAA	TTG	GAA	AGG	ACA	TGT	GAA	CTT	ATA	TAT	TTG	ACA	CAA	CCC
Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	Pro
15					20					25					30
AGC	AGT	TCG	ATA	TCT	ACT	GAA	ATA	AAT	TCT	GCA	TTG	GTG	CTA	AAA	GTT
Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys	Val
				35					40					45	
TCT	TGG	ATC	ACA	TTT	TTA	TTA	GCT	AAA	GGG	GAA	GTA	TTA	CAA	ATG	GAA
Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	Met	Glu
			50					55					60		
GAT	GAT	CTG	GTG	ATT	TCA	TTT	CAG	TTA	ATG	CTA	TGT	GTC	CTT	GAC	TAT
Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp	Tyr
		65					70					75			
TTT	ATT	AAA	CTC	TCA	CCT	CCC	ATG	TTG	CTC	AAA	GAA	CCA	TAT	AAA	ACA
Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys	Thr
	80					85					90				
GCT	GTT	ATA	CCC	ATT	AAT	GGT	TCA	CCT	CGA	ACA	CCC	AGG	CGA	GGT	CAG
Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly	Gln
95					100					105					110
AAC	AGG	AGT	GCA	CGG	ATA	GCA	AAA	CAA	CTA	GAA	AAT	GAT	ACA	AGA	ATT
Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	Arg	Ile
				115					120					125	
ATT	GAA	GTT	CTC	TGT	AAA	GAA	CAT	GAA	TGT	AAT	ATA	GAT	GAG	GTG	AAA
Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu	Val	Lys
			130					135					140		
AAT	GTT	TAT	TTC	AAA	AAT	TTT	ATA	CCT	TTT	ATG	AAT	TCT	CTT	GGA	CTT
Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu	Gly	Leu
		145					150					155			
GTA	ACA	TCT	AAT	GGA	CTT	CCA	GAG	GTT	GAA	AAT	CTT	TCT	AAA	CGA	TAC
Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	Lys	Arg	Tyr
	160					165					170				
GAA	GAA	ATT	TAT	CTT	AAA	AAT	AAA	GAT	CTA	GAT	GCA	AGA	TTA	TTT	TTG
Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	Leu	Phe	Leu
175					180					185					190
GAT	CAT	GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	ATA	GAC	AGT	TTT	GAA	ACA
Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	Phe	Glu	Thr
				195					200					205	

CAG	AGA	ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	GAA	GAG	GTG	AAT	GTA	ATT	672
Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	Asn	Val	Ile	
			210					215					220			
CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	AAC	ACT	ATC	CAA	CAA	TTA	720
Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln	Leu	
			225				230					235				
ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	CCT	TCA	GAA	AAT	CTG	ATT	768
Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu	Ile	
			240				245				250					
TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	AGT	ATA	CTG	AAA	816
Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu	Lys	
255					260					265					270	
AGA	GTG	AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	GAG	AAA	TTT	GCT	AAA	GCT	864
Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys	Ala	
				275					280					285		
GTG	GGA	CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	CAG	CGA	TAC	AAA	CTT	GGA	912
Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	Gly	
			290					295					300			
GTT	CGC	TTG	TAT	TAC	CGA	GTA	ATG	GAA	TCC	ATG	CTT	AAA	TCA	GAA	GAA	960
Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu	Glu	
			305				310					315				
GAA	CGA	TTA	TCC	ATT	CAA	AAT	TTT	AGC	AAA	CTT	CTG	AAT	GAC	AAC	ATT	1008
Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	Asn	Ile	
			320				325				330					
TTT	CAT	ATG	TCT	TTA	TTG	GCG	TGC	GCT	CTT	GAG	GTT	GTA	ATG	GCC	ACA	1056
Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	Thr	
335					340					345					350	
TAT	AGC	AGA	AGT	ACA	TCT	CAG	AAT	CTT	GAT	TCT	GGA	ACA	GAT	TTG	TCT	1104
Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	Ser	
				355					360					365		
TTC	CCA	TGG	ATT	CTG	AAT	GTG	CTT	AAT	TTA	AAA	GCC	TTT	GAT	TTT	TAC	1152
Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	Tyr	
			370					375					380			
AAA	GTG	ATC	GAA	AGT	TTT	ATC	AAA	GCA	GAA	GGC	AAC	TTG	ACA	AGA	GAA	1200
Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	Glu	
			385				390					395				
ATG	ATA	AAA	CAT	TTA	GAA	CGA	TGT	GAA	CAT	CGA	ATC	ATG	GAA	TCC	CTT	1248
Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	Leu	
			400				405				410					

GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA CAA TCA AAG	1296
Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys	
415 420 425 430	
GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT CCT CTT AAT	1344
Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn	
435 440 445	
CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT CTT TCT CCT	1392
Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro	
450 455 460	
GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT	1440
Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr	
465 470 475	
GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA	1488
Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro	
480 485 490	
TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA	1536
Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu	
495 500 505 510	
GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG TCT GAG CAC	1584
Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His	
515 520 525	
CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC ACC CTG CAG	1632
Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln	
530 535 540	
AAT GAG TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA ATT ATG ATG	1680
Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met Met	
545 550 555	
TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC CTT AAA TTC	1728
Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys Phe	
560 565 570	
AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT GTT CAG GAG	1776
Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln Glu	
575 580 585 590	
ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT TCT ATT ATA	1824
Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile	
595 600 605	
GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA AAT ATT TTG	1872
Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile Leu	
610 615 620	

CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA CCT CAC ATT	1920
Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His Ile	
625 630 635	
CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG ATT CCT GGA	1968
Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly	
640 645 650	
GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA ATT TCA GAA	2016
Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu	
655 660 665 670	
GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA ATC TTA GTA	2064
Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val	
675 680 685	
TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG AAA ATA AAT	2112
Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn	
690 695 700	
CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT GCT GAA GGA	2160
Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly	
705 710 715	
AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT ATT GAA GGA	2208
Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly	
720 725 730	
TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG TCC AAA TTT	2256
Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe	
735 740 745 750	
CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA ATG CAA AAG	2304
Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys	
755 760 765	
CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG AAA	2349
Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys	
770 775 780	
TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG TCTCTCACAG	2409
ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT	2469
TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG	2529
TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA TTTAAAAAGT	2589
TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG	2649
GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG	2709
CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC	2769

ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTTGA TTGACTGCCC 2829
 ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA 2889
 AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT ACTGTGTGCT 2949
 TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA 3009
 TACTATCATA CTA CTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT 3069
 TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT 3113

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 781 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala Leu Phe Ser
 1 5 10 15
 Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln Pro Ser Ser
 20 25 30
 Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val Ser Trp
 35 40 45
 Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met Glu Asp Asp
 50 55 60
 Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val Leu Asp Tyr Phe Ile
 65 70 75 80
 Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys Thr Ala Val
 85 90 95
 Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg
 100 105 110
 Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu
 115 120 125
 Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val
 130 135 140
 Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr
 145 150 155 160

Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu
 165 170 175

Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His
 180 185 190

Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg
 195 200 205

Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro
 210 215 220

His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met
 225 230 235 240

Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr
 245 250 255

Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val
 260 265 270

Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly
 275 280 285

Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg
 290 295 300

Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg
 305 310 315 320

Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His
 325 330 335

Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser
 340 345 350

Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro
 355 360 365

Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val
 370 375 380

Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile
 385 390 395 400

Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp
 405 410 415

Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg
 420 425 430

Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro
 435 440 445

Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	Pro	Val	Arg		
450						455					460						
Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	Thr	Ala	Asn		
465					470					475					480		
Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	Pro	Leu	Lys		
				485					490					495			
Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	Leu	Ala	Tyr		
			500					505					510				
Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	His	Pro	Glu		
		515					520					525					
Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	Gln	Asn	Glu		
530						535						540					
Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	Met	Cys	Ser		
545					550					555					560		
Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	Phe	Lys	Ile		
				565					570					575			
Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	Glu	Thr	Phe		
			580					585					590				
Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	Ile	Val	Phe		
		595					600					605					
Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	Leu	Gln	Tyr		
610						615					620						
Ala	Ser	Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	Ile	Pro	Arg		
625					630					635				640			
Ser	Pro	Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	Gly	Gly	Asn		
				645					650					655			
Ile	Tyr	Ile	Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	Glu	Gly	Leu		
			660					665					670				
Pro	Thr	Pro	Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu	Val	Ser	Ile		
		675					680					685					
Gly	Glu	Ser	Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile	Asn	Gln	Met		
690						695					700						
Val	Cys	Asn	Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu	Gly	Ser	Asn		
705					710					715					720		
Pro	Pro	Lys	Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu	Gly	Ser	Asp		
				725					730					735			

Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln
740 745 750

Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys
755 760 765

Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
770 775 780

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3323 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2559

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GTC GAC CTA GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC	144
Val Asp Leu Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn	
35 40 45	
ATA GAA ATC AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT	192
Ile Glu Ile Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp	
50 55 60	
ACC AGT ACC AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT	240
Thr Ser Thr Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr	
65 70 75	
GAT GTA TTG TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT	288
Asp Val Leu Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu	
80 85 90	
ATA TAT TTG ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT	336
Ile Tyr Leu Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser	
95 100 105 110	

GCA	TTG	GTG	CTA	AAA	GTT	TCT	TGG	ATC	ACA	TTT	TTA	TTA	GCT	AAA	GGG	384
Ala	Leu	Val	Leu	Lys	Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	
				115					120					125		
GAA	GTA	TTA	CAA	ATG	GAA	GAT	GAT	CTG	GTG	ATT	TCA	TTT	CAG	TTA	ATG	432
Glu	Val	Leu	Gln	Met	Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	
			130					135					140			
CTA	TGT	GTC	CTT	GAC	TAT	TTT	ATT	AAA	CTC	TCA	CCT	CCC	ATG	TTG	CTC	480
Leu	Cys	Val	Leu	Asp	Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	
		145					150					155				
AAA	GAA	CCA	TAT	AAA	ACA	GCT	GTT	ATA	CCC	ATT	AAT	GGT	TCA	CCT	CGA	528
Lys	Glu	Pro	Tyr	Lys	Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	
	160					165				170						
ACA	CCC	AGG	CGA	GGT	CAG	AAC	AGG	AGT	GCA	CGG	ATA	GCA	AAA	CAA	CTA	576
Thr	Pro	Arg	Arg	Gly	Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	
175					180					185					190	
GAA	AAT	GAT	ACA	AGA	ATT	ATT	GAA	GTT	CTC	TGT	AAA	GAA	CAT	GAA	TGT	624
Glu	Asn	Asp	Thr	Arg	Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	
				195					200					205		
AAT	ATA	GAT	GAG	GTG	AAA	AAT	GTT	TAT	TTC	AAA	AAT	TTT	ATA	CCT	TTT	672
Asn	Ile	Asp	Glu	Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	
			210					215					220			
ATG	AAT	TCT	CTT	GGA	CTT	GTA	ACA	TCT	AAT	GGA	CTT	CCA	GAG	GTT	GAA	720
Met	Asn	Ser	Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	
		225					230					235				
AAT	CTT	TCT	AAA	CGA	TAC	GAA	GAA	ATT	TAT	CTT	AAA	AAT	AAA	GAT	CTA	768
Asn	Leu	Ser	Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	
	240					245					250					
GAT	GCA	AGA	TTA	TTT	TTG	GAT	CAT	GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	816
Asp	Ala	Arg	Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	
255					260					265					270	
ATA	GAC	AGT	TTT	GAA	ACA	CAG	AGA	ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	864
Ile	Asp	Ser	Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	
				275					280					285		
GAA	GAG	GTG	AAT	GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	912
Glu	Glu	Val	Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	
			290					295					300			
AAC	ACT	ATC	CAA	CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	960
Asn	Thr	Ile	Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	
		305					310					315				

CCT TCA GAA AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA	1008
Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro	
320 325 330	
AAA GAA AGT ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA	1056
Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys	
335 340 345 350	
GAG AAA TTT GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA	1104
Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser	
355 360 365	
CAG CGA TAC AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC	1152
Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser	
370 375 380	
ATG CTT AAA TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA	1200
Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys	
385 390 395	
CTT CTG AAT GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT	1248
Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu	
400 405 410	
GAG GTT GTA ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT	1296
Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp	
415 420 425 430	
TCT GGA ACA GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA	1344
Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu	
435 440 445	
AAA GCC TTT GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA	1392
Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu	
450 455 460	
GGC AAC TTG ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT	1440
Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His	
465 470 475	
CGA ATC ATG GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT	1488
Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp	
480 485 490	
CTT ATT AAA CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA	1536
Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu	
495 500 505 510	
TCT GCT TGT CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA	1584
Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala	
515 520 525	

GAT Asp	ATG Met	TAT Tyr	CTT Leu 530	TCT Ser	CCT Pro	GTA Val	AGA Arg	TCT Ser 535	CCA Pro	AAG Lys	AAA Lys	AAA Lys	GGT Gly 540	TCA Ser	ACT Thr	1632
ACG Thr	CGT Arg	GTA Val 545	AAT Asn	TCT Ser	ACT Thr	GCA Ala	AAT Asn 550	GCA Ala	GAG Glu	ACA Thr	CAA Gln 555	GCA Ala	ACC Thr	TCA Ser	GCC Ala	1680
TTC Phe 560	CAG Gln	ACC Thr	CAG Gln	AAG Lys	CCA Pro	TTG Leu 565	AAA Lys	TCT Ser	ACC Thr	TCT Ser	CTT Leu 570	TCA Ser	CTG Leu	TTT Phe	TAT Tyr	1728
AAA Lys 575	AAA Lys	GTG Val	TAT Tyr	CGG Arg 580	CTA Leu	GCC Ala	TAT Tyr 585	CTC Leu	CGG Arg	CTA Leu 585	AAT Asn	ACA Thr	CTT Leu	TGT Cys	GAA Glu 590	1776
CGC Arg	CTT Leu	CTG Leu	TCT Ser 595	GAG Glu	CAC His	CCA Pro	GAA Glu	TTA Leu 600	GAA Glu	CAT His	ATC Ile	ATC Ile	TGG Trp 605	ACC Thr	CTT Leu	1824
TTC Phe	CAG Gln	CAC His 610	ACC Thr	CTG Leu	CAG Gln	AAT Asn	GAG Glu 615	TAT Tyr	GAA Glu	CTC Leu	ATG Met	AGA Arg 620	GAC Asp	AGG Arg	CAT His	1872
TTG Leu	GAC Asp 625	CAA Gln	ATT Ile	ATG Met	ATG Met	TGT Cys 630	TCC Ser	ATG Met	TAT Tyr	GGC Gly	ATA Ile	TGC Cys 635	AAA Lys	GTG Val	AAG Lys	1920
AAT Asn 640	ATA Ile	GAC Asp	CTT Leu	AAA Lys	TTC Phe	AAA Lys 645	ATC Ile	ATT Ile	GTA Val	ACA Thr 650	GCA Ala	TAC Tyr	AAG Lys	GAT Asp	CTT Leu	1968
CCT Pro 655	CAT His	GCT Ala	GTT Val	CAG Gln 660	GAG Glu	ACA Thr	TTC Phe	AAA Lys	CGT Arg 665	GTT Val	TTG Leu	ATC Ile	AAA Lys	GAA Glu	GAG Glu 670	2016
GAG Glu	TAT Tyr	GAT Asp	TCT Ser 675	ATT Ile	ATA Ile	GTA Val	TTC Phe	TAT Tyr 680	AAC Asn	TCG Ser	GTC Val	TTC Phe	ATG Met 685	CAG Gln	AGA Arg	2064
CTG Leu	AAA Lys	ACA Thr 690	AAT Asn	ATT Ile	TTG Leu	CAG Gln	TAT Tyr 695	GCT Ala	TCC Ser	ACC Thr	AGG Arg	CCC Pro 700	CCT Pro	ACC Thr	TTG Leu	2112
TCA Ser	CCA Pro 705	ATA Ile	CCT Pro	CAC His	ATT Ile	CCT Pro	CGA Arg 710	AGC Ser	CCT Pro	TAC Tyr	AAG Lys 715	TTT Phe	CCT Pro	AGT Ser	TCA Ser	2160
CCC Pro 720	TTA Leu	CGG Arg	ATT Ile	CCT Pro	GGA Gly	GGG Gly 725	AAC Asn	ATC Ile	TAT Tyr	ATT Ile	TCA Ser 730	CCC Pro	CTG Leu	AAG Lys	AGT Ser	2208

CCA TAT AAA ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA Pro Tyr Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro 735 740 745 750	2256
AGA TCA AGA ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG Arg Ser Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu 755 760 765	2304
AAG TTC CAG AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu 770 775 780	2352
AAA AGA AGT GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu 785 790 795	2400
CGC TTT GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu 800 805 810	2448
CCA GGA GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr 815 820 825 830	2496
CGA ACA CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser 835 840 845	2544
AAC AAG GAA GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT Asn Lys Glu Glu Lys 850	2599
GGATTCATTG TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC	2659
ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA	2719
TTCCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGCTGT	2779
AATCCTGCCA TTTAAAAAGT TGTCAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG	2839
CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC	2899
TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT	2959
GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAATG TGCCTATCT ATCTTCCAAA	3019
TGCAATTTGA TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT	3079
ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG	3139
AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA	3199
AGCAAAGTAT AACCATATGA TACTATCATA CTAAGTAAAC AGATTTTATA CCTCAGAAATG	3259

TAAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA 3319
TAGT 3323

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 851 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	1	5	10	15
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Val	Asp	20	25	30	
Leu	Asp	Glu	Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	35	40	45	
Ile	Ser	Val	His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	50	55	60	
Thr	Lys	Val	Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	65	70	75	80
Leu	Phe	Ala	Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	85	90	95	
Leu	Thr	Gln	Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	100	105	110	
Val	Leu	Lys	Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	115	120	125	
Leu	Gln	Met	Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	130	135	140	
Val	Leu	Asp	Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	145	150	155	160
Pro	Tyr	Lys	Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	165	170	175	
Arg	Arg	Gly	Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	180	185	190	
Asp	Thr	Arg	Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	195	200	205	

Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala
 500 505 510
 Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met
 515 520 525
 Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg
 530 535 540
 Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln
 545 550 555 560
 Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys
 565 570 575
 Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu
 580 585 590
 Leu Ser Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln
 595 600 605
 His Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp
 610 615 620
 Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile
 625 630 635 640
 Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His
 645 650 655
 Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr
 660 665 670
 Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys
 675 680 685
 Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro
 690 695 700
 Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu
 705 710 715 720
 Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr
 725 730 735
 Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser
 740 745 750
 Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe
 755 760 765
 Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg
 770 775 780

Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe
785 790 795 800

Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly
805 810 815

Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr
820 825 830

Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys
835 840 845

Glu Glu Lys
850

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3461 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2697

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTG GTC GAC	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Val Asp	
65 70 75	
CTA GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA	288
Leu Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu	
80 85 90	

ATC AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT Ile Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser 95 100 105 110	336
ACC AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA Thr Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val 115 120 125	384
TTG TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT Leu Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr 130 135 140	432
TTG ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT GCA TTG Leu Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu 145 150 155	480
GTG CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA Val Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val 160 165 170	528
TTA CAA ATG GAA GAT GAT CTG GTG ATT TCA TTT CAG TTA ATG CTA TGT Leu Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys 175 180 185 190	576
GTC CTT GAC TAT TTT ATT AAA CTC TCA CCT CCC ATG TTG CTC AAA GAA Val Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu 195 200 205	624
CCA TAT AAA ACA GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC Pro Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro 210 215 220	672
AGG CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT Arg Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn 225 230 235	720
GAT ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA Asp Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile 240 245 250	768
GAT GAG GTG AAA AAT GTT TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT Asp Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn 255 260 265 270	816
TCT CTT GGA CTT GTA ACA TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT Ser Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu 275 280 285	864
TCT AAA CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA Ser Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala 290 295 300	912

AGA	TTA	TTT	TTG	GAT	CAT	GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	ATA	GAC	960
Arg	Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	
		305					310					315				
AGT	TTT	GAA	ACA	CAG	AGA	ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	GAA	GAG	1008
Ser	Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	
		320					325				330					
GTG	AAT	GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	AAC	ACT	1056
Val	Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	
		335					340				345				350	
ATC	CAA	CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	CCT	TCA	1104
Ile	Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	
				355					360					365		
GAA	AAT	CTG	ATT	TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	1152
Glu	Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	
		370						375					380			
AGT	ATA	CTG	AAA	AGA	GTG	AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	GAG	AAA	1200
Ser	Ile	Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	
		385					390					395				
TTT	GCT	AAA	GCT	GTG	GGA	CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	CAG	CGA	1248
Phe	Ala	Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	
		400					405				410					
TAC	AAA	CTT	GGA	GTT	CGC	TTG	TAT	TAC	CGA	GTA	ATG	GAA	TCC	ATG	CTT	1296
Tyr	Lys	Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	
		415				420				425					430	
AAA	TCA	GAA	GAA	GAA	CGA	TTA	TCC	ATT	CAA	AAT	TTT	AGC	AAA	CTT	CTG	1344
Lys	Ser	Glu	Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	
				435					440					445		
AAT	GAC	AAC	ATT	TTT	CAT	ATG	TCT	TTA	TTG	GCG	TGC	GCT	CTT	GAG	GTT	1392
Asn	Asp	Asn	Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	
			450					455					460			
GTA	ATG	GCC	ACA	TAT	AGC	AGA	AGT	ACA	TCT	CAG	AAT	CTT	GAT	TCT	GGA	1440
Val	Met	Ala	Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	
		465					470					475				
ACA	GAT	TTG	TCT	TTC	CCA	TGG	ATT	CTG	AAT	GTG	CTT	AAT	TTA	AAA	GCC	1488
Thr	Asp	Leu	Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	
		480					485				490					
TTT	GAT	TTT	TAC	AAA	GTG	ATC	GAA	AGT	TTT	ATC	AAA	GCA	GAA	GGC	AAC	1536
Phe	Asp	Phe	Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	
		495				500				505					510	

TTG	ACA	AGA	GAA	ATG	ATA	AAA	CAT	TTA	GAA	CGA	TGT	GAA	CAT	CGA	ATC	1584
Leu	Thr	Arg	Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	
			515						520					525		
ATG	GAA	TCC	CTT	GCA	TGG	CTC	TCA	GAT	TCA	CCT	TTA	TTT	GAT	CTT	ATT	1632
Met	Glu	Ser	Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	
			530					535					540			
AAA	CAA	TCA	AAG	GAC	CGA	GAA	GGA	CCA	ACT	GAT	CAC	CTT	GAA	TCT	GCT	1680
Lys	Gln	Ser	Lys	Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	
			545				550					555				
TGT	CCT	CTT	AAT	CTT	CCT	CTC	CAG	AAT	AAT	CAC	ACT	GCA	GCA	GAT	ATG	1728
Cys	Pro	Leu	Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	
	560					565					570					
TAT	CTT	TCT	CCT	GTA	AGA	TCT	CCA	AAG	AAA	AAA	GGT	TCA	ACT	ACG	CGT	1776
Tyr	Leu	Ser	Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	
575					580				585						590	
GTA	AAT	TCT	ACT	GCA	AAT	GCA	GAG	ACA	CAA	GCA	ACC	TCA	GCC	TTC	CAG	1824
Val	Asn	Ser	Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	
				595				600						605		
ACC	CAG	AAG	CCA	TTG	AAA	TCT	ACC	TCT	CTT	TCA	CTG	TTT	TAT	AAA	AAA	1872
Thr	Gln	Lys	Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	
			610					615					620			
GTG	TAT	CGG	CTA	GCC	TAT	CTC	CGG	CTA	AAT	ACA	CTT	TGT	GAA	CGC	CTT	1920
Val	Tyr	Arg	Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	
		625				630						635				
CTG	TCT	GAG	CAC	CCA	GAA	TTA	GAA	CAT	ATC	ATC	TGG	ACC	CTT	TTC	CAG	1968
Leu	Ser	Glu	His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	
	640					645					650					
CAC	ACC	CTG	CAG	AAT	GAG	TAT	GAA	CTC	ATG	AGA	GAC	AGG	CAT	TTG	GAC	2016
His	Thr	Leu	Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	
655					660				665					670		
CAA	ATT	ATG	ATG	TGT	TCC	ATG	TAT	GGC	ATA	TGC	AAA	GTG	AAG	AAT	ATA	2064
Gln	Ile	Met	Met	Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	
				675				680					685			
GAC	CTT	AAA	TTC	AAA	ATC	ATT	GTA	ACA	GCA	TAC	AAG	GAT	CTT	CCT	CAT	2112
Asp	Leu	Lys	Phe	Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	
			690					695					700			
GCT	GTT	CAG	GAG	ACA	TTC	AAA	CGT	GTT	TTG	ATC	AAA	GAA	GAG	GAG	TAT	2160
Ala	Val	Gln	Glu	Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	
		705					710					715				

GAT TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA	2208
Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys	
720 725 730	
ACA AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA	2256
Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro	
735 740 745 750	
ATA CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA	2304
Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu	
755 760 765	
CGG ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT	2352
Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr	
770 775 780	
AAA ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA	2400
Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser	
785 790 795	
AGA ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC	2448
Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe	
800 805 810	
CAG AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA	2496
Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg	
815 820 825 830	
AGT GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT	2544
Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe	
835 840 845	
GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA	2592
Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly	
850 855 860	
GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA	2640
Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr	
865 870 875	
CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG	2688
Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys	
880 885 890	
GAA GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT	2737
Glu Glu Lys	
895	
GGATTCATTG TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC	2797
ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAATGTGC AGATGCAATT GTTTGGGTGA	2857
TTCCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA	2917

AATCCTGCCA TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG 2977

CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC 3037

TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT 3097

GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCTTATCT ATCTTCCAAA 3157

TGCAATTTGA TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT 3217

ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG 3277

AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA 3337

AGCAAAGTAT AACCATATGA TACTATCATA CTA CTGAAAC AGATTTTCATA CCTCAGAATG 3397

TAAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA 3457

TAGT 3461

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 897 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	1	5	10	15
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp	20	25	30	
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	35	40	45	
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	50	55	60	
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Val	Asp	Leu	Asp	65	70	75	80
Glu	Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	85	90	95	
Val	His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	100	105	110	

Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	
			405						410					415		
Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	
			420					425					430			
Glu	Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	
		435					440					445				
Asn	Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	
	450					455					460					
Ala	Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	
465					470					475					480	
Leu	Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	
			485						490					495		
Phe	Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	
			500					505					510			
Arg	Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	
		515					520					525				
Ser	Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	
	530					535					540					
Ser	Lys	Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	
545					550					555					560	
Leu	Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	
				565				570						575		
Ser	Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	
			580					585					590			
Ser	Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	
		595					600					605				
Lys	Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	
	610					615					620					
Arg	Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	
625					630					635					640	
Glu	His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	
				645					650					655		
Leu	Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	
		660						665					670			
Met	Met	Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	
		675					680					685				

Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val
690 695 700

Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser
705 710 715 720

Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn
725 730 735

Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro
740 745 750

His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile
755 760 765

Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile
770 775 780

Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile
785 790 795 800

Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys
805 810 815

Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala
820 825 830

Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile
835 840 845

Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser
850 855 860

Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met
865 870 875 880

Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu
885 890 895

Lys

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3347 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2583

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GTC GAA TCT ACT GAA ATA AAT TCT GCA TTG GTG CTA AAA GTT TCT TGG	384
Val Glu Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val Ser Trp	
115 120 125	
ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA CAA ATG GAA GAT GAT	432
Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met Glu Asp Asp	
130 135 140	
CTG GTG ATT TCA TTT CAG TTA ATG CTA TGT GTC CTT GAC TAT TTT ATT	480
Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val Leu Asp Tyr Phe Ile	
145 150 155	
AAA CTC TCA CCT CCC ATG TTG CTC AAA GAA CCA TAT AAA ACA GCT GTT	528
Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys Thr Ala Val	
160 165 170	
ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG CGA GGT CAG AAC AGG	576
Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg	
175 180 185 190	

AGT	GCA	CGG	ATA	GCA	AAA	CAA	CTA	GAA	AAT	GAT	ACA	AGA	ATT	ATT	GAA	624
Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	Arg	Ile	Ile	Glu	
				195					200					205		
GTT	CTC	TGT	AAA	GAA	CAT	GAA	TGT	AAT	ATA	GAT	GAG	GTG	AAA	AAT	GTT	672
Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu	Val	Lys	Asn	Val	
			210					215					220			
TAT	TTC	AAA	AAT	TTT	ATA	CCT	TTT	ATG	AAT	TCT	CTT	GGA	CTT	GTA	ACA	720
Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu	Gly	Leu	Val	Thr	
		225					230					235				
TCT	AAT	GGA	CTT	CCA	GAG	GTT	GAA	AAT	CTT	TCT	AAA	CGA	TAC	GAA	GAA	768
Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	Lys	Arg	Tyr	Glu	Glu	
	240					245					250					
ATT	TAT	CTT	AAA	AAT	AAA	GAT	CTA	GAT	GCA	AGA	TTA	TTT	TTG	GAT	CAT	816
Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	Leu	Phe	Leu	Asp	His	
255					260					265					270	
GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	ATA	GAC	AGT	TTT	GAA	ACA	CAG	AGA	864
Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	Phe	Glu	Thr	Gln	Arg	
				275					280					285		
ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	GAA	GAG	GTG	AAT	GTA	ATT	CCT	CCA	912
Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	Asn	Val	Ile	Pro	Pro	
			290					295					300			
CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	AAC	ACT	ATC	CAA	CAA	TTA	ATG	ATG	960
His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln	Leu	Met	Met	
		305					310					315				
ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	CCT	TCA	GAA	AAT	CTG	ATT	TCC	TAT	1008
Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu	Ile	Ser	Tyr	
	320					325					330					
TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	AGT	ATA	CTG	AAA	AGA	GTG	1056
Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu	Lys	Arg	Val	
335					340					345					350	
AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	GAG	AAA	TTT	GCT	AAA	GCT	GTG	GGA	1104
Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys	Ala	Val	Gly	
				355					360					365		
CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	CAG	CGA	TAC	AAA	CTT	GGA	GTT	CGC	1152
Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	Gly	Val	Arg	
			370					375					380			
TTG	TAT	TAC	CGA	GTA	ATG	GAA	TCC	ATG	CTT	AAA	TCA	GAA	GAA	GAA	CGA	1200
Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu	Glu	Glu	Arg	
		385					390					395				

TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT GAC AAC ATT TTT CAT	1248
Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His	
400 405 410	
ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA ATG GCC ACA TAT AGC	1296
Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser	
415 420 425 430	
AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA GAT TTG TCT TTC CCA	1344
Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro	
435 440 445	
TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT GAT TTT TAC AAA GTG	1392
Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val	
450 455 460	
ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG ACA AGA GAA ATG ATA	1440
Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile	
465 470 475	
AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG GAA TCC CTT GCA TGG	1488
Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp	
480 485 490	
CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA CAA TCA AAG GAC CGA	1536
Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg	
495 500 505 510	
GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT CCT CTT AAT CTT CCT	1584
Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro	
515 520 525	
CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT CTT TCT CCT GTA AGA	1632
Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg	
530 535 540	
TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT GCA AAT	1680
Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn	
545 550 555	
GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA TTG AAA	1728
Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys	
560 565 570	
TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA GCC TAT	1776
Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr	
575 580 585 590	
CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG TCT GAG CAC CCA GAA	1824
Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His Pro Glu	
595 600 605	

TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC ACC CTG CAG AAT GAG Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln Asn Glu 610 615 620	1872
TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA ATT ATG ATG TGT TCC Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met Met Cys Ser 625 630 635	1920
ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC CTT AAA TTC AAA ATC Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys Phe Lys Ile 640 645 650	1968
ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT GTT CAG GAG ACA TTC Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln Glu Thr Phe 655 660 665 670	2016
AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT TCT ATT ATA GTA TTC Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe 675 680 685	2064
TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA AAT ATT TTG CAG TAT Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr 690 695 700	2112
GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA CCT CAC ATT CCT CGA Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His Ile Pro Arg 705 710 715	2160
AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG ATT CCT GGA GGG AAC Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn 720 725 730	2208
ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA ATT TCA GAA GGT CTG Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu 735 740 745 750	2256
CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA ATC TTA GTA TCA ATT Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val Ser Ile 755 760 765	2304
GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG AAA ATA AAT CAG ATG Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn Gln Met 770 775 780	2352
GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT GCT GAA GGA AGC AAC Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly Ser Asn 785 790 795	2400
CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT ATT GAA GGA TCA GAT Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp 800 805 810	2448

GAA GCA GAT	GGA AGT AAA CAT CTC CCA GGA GAG TCC AAA TTT CAG CAG	
Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln		2496
815 820 825 830		
AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA ATG CAA AAG CAG AAA		2544
Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys		
835 840 845		
ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG AAA TGAGGATCTC		2593
Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys		
850 855		
AGGACCTTG	TGGACTGT GTACACCTCT GGATTTCATTG TCTCTCACAG ATGTGACTGT	2653
ATAACTTTCC	CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT TTTGTGGATA	2713
TAAAAATGTGC	AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG TTAGTCATTG	2773
TTATTTATAC	AAGATTGAAA ATCTTGTGTA AATCCTGCCA TTTAAAAAGT TGTAGCAGAT	2833
TGTTTTCTCT	TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG GCCCTAGAGT	2893
GGGAGTCCTG	ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG CATATAGGTG	2953
ATGTTTGCTC	TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC ATGAACACCC	3013
TTAGAAAATG	TGTCCTATCT ATCTTCCAAA TGCAATTTGA TTGACTGCCC ATTCACCAA	3073
ATTATCCTGA	ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA AATTACTAAT	3133
TTTACACATT	AGATTTTATT TTACTATTGG AATCTGATAT ACTGTGTGCT TGT'TTTATAA	3193
AATTTTGCTT	TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA TACTATCAT	3253
CTACTGAAAC	AGATTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT TTTCTTCATC	3313
CAACTTATGT	TTTTAAATGA GGATTATTGA TAGT	3347

(2) INFORMATION FOR SEQ ID NO:43:

(ii) MOLECULE TYPE: protein

Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala Ala Ala
1 5 10 15

Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Pro Glu Glu Asp
20 25 30

Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg Leu Glu
35 40 45

Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln Lys Leu
50 55 60

Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp Glu Lys
65 70 75 80

Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys Lys Lys
85 90 95

Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Val Glu
100 105 110

Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val Ser Trp Ile Thr
115 120 125

Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met Glu Asp Asp Leu Val
130 135 140

Ile Ser Phe Gln Leu Met Leu Cys Val Leu Asp Tyr Phe Ile Lys Leu
145 150 155 160

Ser Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys Thr Ala Val Ile Pro
165 170 175

Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg Ser Ala
180 185 190

Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu Val Leu
195 200 205

Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val Tyr Phe
210 215 220

Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr Ser Asn
225 230 235 240

Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile Tyr
245 250 255

Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp Lys
260 265 270

Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro
275 280 285

Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr
290 295 300

Pro 305	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln	Leu	Met	Met	Ile	Leu
				310				315				320			
Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu	Ile	Ser	Tyr	Phe	Asn
				325				330				335			
Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu	Lys	Arg	Val	Lys	Asp
				340				345				350			
Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys	Ala	Val	Gly	Gln	Gly
				355				360				365			
Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	Gly	Val	Arg	Leu	Tyr
				370				375				380			
Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu	Glu	Glu	Arg	Leu	Ser
385				390				395				400			
Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	Asn	Ile	Phe	His	Met	Ser
				405				410				415			
Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	Thr	Tyr	Ser	Arg	Ser
				420				425				430			
Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	Ser	Phe	Pro	Trp	Ile
				435				440				445			
Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	Tyr	Lys	Val	Ile	Glu
				450				455				460			
Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	Glu	Met	Ile	Lys	His
465				470				475				480			
Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	Leu	Ala	Trp	Leu	Ser
				485				490				495			
Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	Lys	Asp	Arg	Glu	Gly
				500				505				510			
Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	Asn	Leu	Pro	Leu	Gln
				515				520				525			
Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	Pro	Val	Arg	Ser	Pro
530				535				540				545			
Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	Thr	Ala	Asn	Ala	Glu
545				550				555				560			
Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	Pro	Leu	Lys	Ser	Thr
				565				570				575			
Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	Leu	Ala	Tyr	Leu	Arg
				580				585				590			

Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	His	Pro	Glu	Leu	Glu
595						600						605			
His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	Gln	Asn	Glu	Tyr	Glu
610						615						620			
Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	Met	Cys	Ser	Met	Tyr
625						630						635			
Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	Phe	Lys	Ile	Ile	Val
			645						650						
Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	Glu	Thr	Phe	Lys	Arg
			660						665						
Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	Ile	Val	Phe	Tyr	Asn
675						680						685			
Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	Leu	Gln	Tyr	Ala	Ser
690						695						700			
Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	Ile	Pro	Arg	Ser	Pro
705						710						715			
Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	Gly	Gly	Asn	Ile	Tyr
			725						730						
Ile	Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	Glu	Gly	Leu	Pro	Thr
			740						745						
Pro	Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu	Val	Ser	Ile	Gly	Glu
755						760						765			
Ser	Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile	Asn	Gln	Met	Val	Cys
770						775						780			
Asn	Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu	Gly	Ser	Asn	Pro	Pro
785						790						795			
Lys	Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu	Gly	Ser	Asp	Glu	Ala
			805						810						
Asp	Gly	Ser	Lys	His	Leu	Pro	Gly	Glu	Ser	Lys	Phe	Gln	Gln	Lys	Leu
			820						825						
Ala	Glu	Met	Thr	Ser	Thr	Arg	Thr	Arg	Met	Gln	Lys	Gln	Lys	Met	Asn
835						840						845			
Asp	Ser	Met	Asp	Thr	Ser	Asn	Lys	Glu	Glu	Lys					
850						855									

(2) INFORMATION FOR SEO ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3161 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 7..2397

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG CGA GGT CAG	384
Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln	
115 120 125	
AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT ACA AGA ATT	432
Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile	
130 135 140	
ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT GAG GTG AAA	480
Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys	
145 150 155	

AAT GTT TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT CTT GGA CTT Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu 160 165 170	528
GTA ACA TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT AAA CGA TAC Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr 175 180 185 190	576
GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA TTA TTT TTG Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu 195 200 205	624
GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT TTT GAA ACA Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr 210 215 220	672
CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG AAT GTA ATT Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile 225 230 235	720
CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC CAA CAA TTA Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu 240 245 250	768
ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA AAT CTG ATT Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile 255 260 265 270	816
TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT ATA CTG AAA Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys 275 280 285	864
AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT GCT AAA GCT Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala 290 295 300	912
GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC AAA CTT GGA Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly 305 310 315	960
GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA TCA GAA GAA Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu 320 325 330	1008
GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT GAC AAC ATT Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile 335 340 345 350	1056
TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA ATG GCC ACA Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr 355 360 365	1104

TAT	AGC	AGA	AGT	ACA	TCT	CAG	AAT	CTT	GAT	TCT	GGA	ACA	GAT	TTG	TCT	1152
Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	Ser	
			370					375					380			
TTC	CCA	TGG	ATT	CTG	AAT	GTG	CTT	AAT	TTA	AAA	GCC	TTT	GAT	TTT	TAC	1200
Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	Tyr	
		385					390					395				
AAA	GTG	ATC	GAA	AGT	TTT	ATC	AAA	GCA	GAA	GGC	AAC	TTG	ACA	AGA	GAA	1248
Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	Glu	
	400					405					410					
ATG	ATA	AAA	CAT	TTA	GAA	CGA	TGT	GAA	CAT	CGA	ATC	ATG	GAA	TCC	CTT	1296
Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	Leu	
415					420				425						430	
GCA	TGG	CTC	TCA	GAT	TCA	CCT	TTA	TTT	GAT	CTT	ATT	AAA	CAA	TCA	AAG	1344
Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	Lys	
			435						440				445			
GAC	CGA	GAA	GGA	CCA	ACT	GAT	CAC	CTT	GAA	TCT	GCT	TGT	CCT	CTT	AAT	1392
Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	Asn	
		450					455					460				
CTT	CCT	CTC	CAG	AAT	AAT	CAC	ACT	GCA	GCA	GAT	ATG	TAT	CTT	TCT	CCT	1440
Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	Pro	
		465				470						475				
GTA	AGA	TCT	CCA	AAG	AAA	AAA	GGT	TCA	ACT	ACG	CGT	GTA	AAT	TCT	ACT	1488
Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	Thr	
	480					485					490					
GCA	AAT	GCA	GAG	ACA	CAA	GCA	ACC	TCA	GCC	TTC	CAG	ACC	CAG	AAG	CCA	1536
Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	Pro	
495					500				505						510	
TTG	AAA	TCT	ACC	TCT	CTT	TCA	CTG	TTT	TAT	AAA	AAA	GTG	TAT	CGG	CTA	1584
Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	Leu	
			515					520					525			
GCC	TAT	CTC	CGG	CTA	AAT	ACA	CTT	TGT	GAA	CGC	CTT	CTG	TCT	GAG	CAC	1632
Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	His	
		530					535						540			
CCA	GAA	TTA	GAA	CAT	ATC	ATC	TGG	ACC	CTT	TTC	CAG	CAC	ACC	CTG	CAG	1680
Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	Gln	
		545					550					555				
AAT	GAG	TAT	GAA	CTC	ATG	AGA	GAC	AGG	CAT	TTG	GAC	CAA	ATT	ATG	ATG	1728
Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	Met	
	560					565				570						

TGT Cys 575	TCC Ser	ATG Met	TAT Tyr	GGC Gly	ATA Ile 580	TGC Cys	AAA Lys	GTG Val	AAG Lys	AAT Asn 585	ATA Ile	GAC Asp	CTT Leu	AAA Lys	TTC Phe 590	1776
AAA Lys	ATC Ile	ATT Ile	GTA Val 595	ACA Thr	GCA Ala	TAC Tyr	AAG Lys	GAT Asp	CTT Leu 600	CCT Pro	CAT His	GCT Ala	GTT Val 605	CAG Gln	GAG Glu	1824
ACA Thr	TTC Phe	AAA Lys	CGT Arg 610	GTT Val	TTG Leu	ATC Ile	AAA Lys	GAA Glu 615	GAG Glu	GAG Glu	TAT Tyr	GAT Asp	TCT Ser 620	ATT Ile	ATA Ile	1872
GTA Val	TTC Phe 625	TAT Tyr	AAC Asn	TCG Ser	GTC Val	TTC Phe 630	ATG Met	CAG Gln	AGA Arg	CTG Leu	AAA Lys 635	ACA Thr	AAT Asn	ATT Ile	TTG Leu	1920
CAG Gln 640	TAT Tyr	GCT Ala	TCC Ser	ACC Thr	AGG Arg 645	CCC Pro	CCT Pro	ACC Thr	TTG Leu	TCA Ser 650	CCA Pro	ATA Ile	CCT Pro	CAC His	ATT Ile	1968
CCT Pro 655	CGA Arg	AGC Ser	CCT Pro	TAC Tyr 660	AAG Lys	TTT Phe	CCT Pro	AGT Ser	TCA Ser 665	CCC Pro	TTA Leu	CGG Arg	ATT Ile	CCT Pro	GGA Gly 670	2016
GGG Gly	AAC Asn	ATC Ile	TAT Tyr 675	ATT Ile	TCA Ser	CCC Pro	CTG Leu	AAG Lys 680	AGT Ser	CCA Pro	TAT Tyr	AAA Lys	ATT Ile 685	TCA Ser	GAA Glu	2064
GGT Gly	CTG Leu	CCA Pro 690	ACA Thr	CCA Pro	ACA Thr	AAA Lys	ATG Met 695	ACT Thr	CCA Pro	AGA Arg	TCA Ser	AGA Arg	ATC Ile 700	TTA Leu	GTA Val	2112
TCA Ser	ATT Ile 705	GGT Gly	GAA Glu	TCA Ser	TTC Phe	GGG Gly 710	ACT Thr	TCT Ser	GAG Glu	AAG Lys	TTC Phe 715	CAG Gln	AAA Lys	ATA Ile	AAT Asn	2160
CAG Gln 720	ATG Met	GTA Val	TGT Cys	AAC Asn	AGC Ser	GAC Asp 725	CGT Arg	GTG Val	CTC Leu	AAA Lys 730	AGA Arg	AGT Ser	GCT Ala	GAA Glu	GGA Gly	2208
AGC Ser 735	AAC Asn	CCT Pro	CCT Pro	AAA Lys 740	CCA Pro	CTG Leu	AAA Lys	AAA Lys	CTA Leu 745	CGC Arg	TTT Phe	GAT Asp	ATT Ile	GAA Glu 750	GGA Gly	2256
TCA Ser	GAT Asp	GAA Glu	GCA Ala 755	GAT Asp	GGA Gly	AGT Ser	AAA Lys	CAT His 760	CTC Leu	CCA Pro	GGA Gly	GAG Glu	TCC Ser 765	AAA Lys	TTT Phe	2304
CAG Gln	CAG Gln	AAA Lys	CTG Leu 770	GCA Ala	GAA Glu	ATG Met	ACT Thr 775	TCT Ser	ACT Thr	CGA Arg	ACA Thr	CGA Arg	ATG Met 780	CAA Gln	AAG Lys	2352

CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG AAA	2397
Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys	
785 790 795	
TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG TCTCTCACAG	2457
ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT	2517
TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG	2577
TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGCTGTA AATCCTGCCA TTTAAAAAGT	2637
TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG	2697
GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG	2757
CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC	2817
ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTTGA TTGACTGCCC	2877
ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA	2937
AATTACTAAT TTACACATT AGATTTTATT TTACTATTGG AATCTGATAT ACTGTGTGCT	2997
TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA	3057
TACTATCATA CTACTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT	3117
TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT	3161

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 797 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala Ala Ala	
1 5 10 15	
Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu Glu Asp	
20 25 30	
Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg Leu Glu	
35 40 45	
Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln Lys Leu	
50 55 60	

Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser
 355 360 365

Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro
 370 375 380

Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val
 385 390 395 400

Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile
 405 410 415

Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp
 420 425 430

Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg
 435 440 445

Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro
 450 455 460

Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg
 465 470 475 480

Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn
 485 490 495

Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys
 500 505 510

Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr
 515 520 525

Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His Pro Glu
 530 535 540

Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln Asn Glu
 545 550 555 560

Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met Met Cys Ser
 565 570 575

Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys Phe Lys Ile
 580 585 590

Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln Glu Thr Phe
 595 600 605

Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe
 610 615 620

Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr
 625 630 635 640

Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His Ile Pro Arg
645 650 655

Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn
660 665 670

Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu
675 680 685

Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val Ser Ile
690 695 700

Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn Gln Met
705 710 715 720

Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly Ser Asn
725 730 735

Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp
740 745 750

Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln
755 760 765

Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys
770 775 780

Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
785 790 795

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3377 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2613

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	

GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	384
Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
115 120 125	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	432
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
130 135 140	
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG	480
Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu	
145 150 155	
TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG	528
Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu	
160 165 170	
ACA CAA CCC AGC AGT TCG ATG GTC GCT GTT ATA CCC ATT AAT GGT TCA	576
Thr Gln Pro Ser Ser Ser Met Val Ala Val Ile Pro Ile Asn Gly Ser	
175 180 185 190	
CCT CGA ACA CCC AGG CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA	624
Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys	
195 200 205	
CAA CTA GAA AAT GAT ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT	672
Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His	
210 215 220	
GAA TGT AAT ATA GAT GAG GTG AAA AAT GTT TAT TTC AAA AAT TTT ATA	720
Glu Cys Asn Ile Asp Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile	
225 230 235	

CCT TTT ATG AAT TCT CTT GGA CTT GTA ACA TCT AAT GGA CTT CCA GAG	768
Pro Phe Met Asn Ser Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu	
240 245 250	
GTT GAA AAT CTT TCT AAA CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA	816
Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys	
255 260 265 270	
GAT CTA GAT GCA AGA TTA TTT TTG GAT CAT GAT AAA ACT CTT CAG ACT	864
Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr	
275 280 285	
GAT TCT ATA GAC AGT TTT GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC	912
Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn	
290 295 300	
CTT GAT GAA GAG GTG AAT GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT	960
Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr Pro Val Arg Thr	
305 310 315	
GTT ATG AAC ACT ATC CAA CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT	1008
Val Met Asn Thr Ile Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser	
320 325 330	
GAT CAA CCT TCA GAA AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG	1056
Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val	
335 340 345 350	
AAT CCA AAA GAA AGT ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC	1104
Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile	
355 360 365	
TTT AAA GAG AAA TTT GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT	1152
Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile	
370 375 380	
GGA TCA CAG CGA TAC AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG	1200
Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met	
385 390 395	
GAA TCC ATG CTT AAA TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT	1248
Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe	
400 405 410	
AGC AAA CTT CTG AAT GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC	1296
Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys	
415 420 425 430	
GCT CTT GAG GTT GTA ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT	1344
Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn	
435 440 445	

CTT GAT TCT GGA ACA GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu 450 455 460	1392
AAT TTA AAA GCC TTT GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys 465 470 475	1440
GCA GAA GGC AAC TTG ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys 480 485 490	1488
GAA CAT CGA ATC ATG GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu 495 500 505 510	1536
TTT GAT CTT ATT AAA CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His 515 520 525	1584
CTT GAA TCT GCT TGT CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr 530 535 540	1632
GCA GCA GAT ATG TAT CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly 545 550 555	1680
TCA ACT ACG CGT GTA AAT TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr 560 565 570	1728
TCA GCC TTC CAG ACC CAG AAG CCA TTG AAA TCT ACC TCT CTT TCA CTG Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu 575 580 585 590	1776
TTT TAT AAA AAA GTG TAT CGG CTA GCC TAT CTC CGG CTA AAT ACA CTT Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu 595 600 605	1824
TGT GAA CGC CTT CTG TCT GAG CAC CCA GAA TTA GAA CAT ATC ATC TGG Cys Glu Arg Leu Leu Ser Glu His Pro Glu Leu Glu His Ile Ile Trp 610 615 620	1872
ACC CTT TTC CAG CAC ACC CTG CAG AAT GAG TAT GAA CTC ATG AGA GAC Thr Leu Phe Gln His Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp 625 630 635	1920
AGG CAT TTG GAC CAA ATT ATG ATG TGT TCC ATG TAT GGC ATA TGC AAA Arg His Leu Asp Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys 640 645 650	1968

GTG AAG AAT ATA GAC CTT AAA TTC AAA ATC ATT GTA ACA GCA TAC AAG	2016
Val Lys Asn Ile Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys	
655 660 665 670	
GAT CTT CCT CAT GCT GTT CAG GAG ACA TTC AAA CGT GTT TTG ATC AAA	2064
Asp Leu Pro His Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys	
675 680 685	
GAA GAG GAG TAT GAT TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG	2112
Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met	
690 695 700	
CAG AGA CTG AAA ACA AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT	2160
Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro	
705 710 715	
ACC TTG TCA CCA ATA CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT	2208
Thr Leu Ser Pro Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro	
720 725 730	
AGT TCA CCC TTA CGG ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG	2256
Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu	
735 740 745 750	
AAG AGT CCA TAT AAA ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG	2304
Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met	
755 760 765	
ACT CCA AGA TCA AGA ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT	2352
Thr Pro Arg Ser Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr	
770 775 780	
TCT GAG AAG TTC CAG AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT	2400
Ser Glu Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg	
785 790 795	
GTG CTC AAA AGA AGT GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA	2448
Val Leu Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys	
800 805 810	
AAA CTA CGC TTT GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA	2496
Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys	
815 820 825 830	
CAT CTC CCA GGA GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT	2544
His Leu Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr	
835 840 845	
TCT ACT CGA ACA CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT	2592
Ser Thr Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp	
850 855 860	

ACC TCA AAC AAG GAA GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT 2643
 Thr Ser Asn Lys Glu Glu Lys
 865

GTACACCTCT GGATTCATTG TCTCTCACAG ATGTGACTGT ATAACTTTCC CAGGTTCTGT 2703
 TTATGGCCAC ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT 2763
 GTTTGGGTGA TTCCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA 2823
 ATCTTGTTGTA AATCCTGCCA TTTAAAAAGT TGTCAGCAGAT TGTTTCCTCT TCCAAAGTAA 2883
 AATTGCTGTG CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG 2943
 CCTGTCTGAC TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT 3003
 TAATTTATAT GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCTTATCT 3063
 ATCTTCCAAA TGCAATTTGA TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC 3123
 AAAAATGGAT ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT 3183
 TTACTATTGG AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT 3243
 AAAAGCTGGA AGCAAAGTAT AACCATATGA TACTATCATA CTACTGAAAC AGATTTTCATA 3303
 CCTCAGAATG TAAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA 3363
 GGATTATTGA TAGT 3377

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 869 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala
1				5					10				15	
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp
			20					25				30		
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu
		35					40				45			
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys
		50				55					60			

Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	65	70	75	80
Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	85	90	95	
Glu	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu	100	105	110	
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	115	120	125	
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	130	135	140	
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	145	150	155	160
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	165	170	175	
Pro	Ser	Ser	Ser	Met	Val	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	180	185	190	
Thr	Pro	Arg	Arg	Gly	Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	195	200	205	
Glu	Asn	Asp	Thr	Arg	Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	210	215	220	
Asn	Ile	Asp	Glu	Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	225	230	235	240
Met	Asn	Ser	Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	245	250	255	
Asn	Leu	Ser	Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	260	265	270	
Asp	Ala	Arg	Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	275	280	285	
Ile	Asp	Ser	Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	290	295	300	
Glu	Glu	Val	Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	305	310	315	320
Asn	Thr	Ile	Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	325	330	335	
Pro	Ser	Glu	Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	340	345	350	

Lys	Glu	Ser	Ile	Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	
	355						360					365				
Glu	Lys	Phe	Ala	Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	
	370					375					380					
Gln	Arg	Tyr	Lys	Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	
385					390					395					400	
Met	Leu	Lys	Ser	Glu	Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	
			405						410					415		
Leu	Leu	Asn	Asp	Asn	Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	
		420						425					430			
Glu	Val	Val	Met	Ala	Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	
	435						440					445				
Ser	Gly	Thr	Asp	Leu	Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	
	450					455					460					
Lys	Ala	Phe	Asp	Phe	Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	
465					470					475					480	
Gly	Asn	Leu	Thr	Arg	Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	
			485					490						495		
Arg	Ile	Met	Glu	Ser	Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	
		500						505					510			
Leu	Ile	Lys	Gln	Ser	Lys	Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	
		515					520					525				
Ser	Ala	Cys	Pro	Leu	Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	
	530					535					540					
Asp	Met	Tyr	Leu	Ser	Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	
545					550					555					560	
Thr	Arg	Val	Asn	Ser	Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	
			565						570					575		
Phe	Gln	Thr	Gln	Lys	Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	
		580						585					590			
Lys	Lys	Val	Tyr	Arg	Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	
	595					600						605				
Arg	Leu	Leu	Ser	Glu	His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	
	610					615					620					
Phe	Gln	His	Thr	Leu	Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	
625					630					635					640	

Leu Asp Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys
 645 650 655
 Asn Ile Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu
 660 665 670
 Pro His Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu
 675 680 685
 Glu Tyr Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg
 690 695 700
 Leu Lys Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu
 705 710 715 720
 Ser Pro Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser
 725 730 735
 Pro Leu Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser
 740 745 750
 Pro Tyr Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro
 755 760 765
 Arg Ser Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu
 770 775 780
 Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu
 785 790 795 800
 Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu
 805 810 815
 Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu
 820 825 830
 Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr
 835 840 845
 Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser
 850 855 860
 Asn Lys Glu Glu Lys
 865

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3383 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 7..2619

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	384
Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
115 120 125	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	432
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
130 135 140	
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG	480
Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu	
145 150 155	
TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG	528
Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu	
160 165 170	
ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT GCA TTG GTG	576
Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val	
175 180 185 190	

CTA	AAA	GTT	TCT	TGG	ATC	ACA	TTT	TTA	TTA	GCT	AAA	GGG	GAA	GTA	TTA	624		
Leu	Lys	Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu			
				195					200				205					
CAA	ATG	GAA	GAT	GAT	CTG	GTG	ATT	TCA	TTT	CAG	TTA	ATG	CTA	TGT	GTC	672		
Gln	Met	Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val			
				210					215				220					
CTT	GAC	TAT	TTT	ATT	AAA	CTC	TCA	CCT	CCC	ATG	TTG	CTC	AAA	GAA	CCA	720		
Leu	Asp	Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro			
				225					230				235					
TAT	AAA	ACA	GGG	TCG	AAT	TCT	CTT	GGA	CTT	GTA	ACA	TCT	AAT	GGA	CTT	768		
Tyr	Lys	Thr	Gly	Ser	Asn	Ser	Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu			
				240					245				250					
CCA	GAG	GTT	GAA	AAT	CTT	TCT	AAA	CGA	TAC	GAA	GAA	ATT	TAT	CTT	AAA	816		
Pro	Glu	Val	Glu	Asn	Leu	Ser	Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys			
				255					260				265				270	
AAT	AAA	GAT	CTA	GAT	GCA	AGA	TTA	TTT	TTG	GAT	CAT	GAT	AAA	ACT	CTT	864		
Asn	Lys	Asp	Leu	Asp	Ala	Arg	Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu			
				275					280				285					
CAG	ACT	GAT	TCT	ATA	GAC	AGT	TTT	GAA	ACA	CAG	AGA	ACA	CCA	CGA	AAA	912		
Gln	Thr	Asp	Ser	Ile	Asp	Ser	Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys			
				290					295				300					
AGT	AAC	CTT	GAT	GAA	GAG	GTG	AAT	GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	960		
Ser	Asn	Leu	Asp	Glu	Glu	Val	Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val			
				305					310				315					
AGG	ACT	GTT	ATG	AAC	ACT	ATC	CAA	CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	1008		
Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser			
				320					325				330					
GCA	AGT	GAT	CAA	CCT	TCA	GAA	AAT	CTG	ATT	TCC	TAT	TTT	AAC	AAC	TGC	1056		
Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys			
				335					340				345				350	
ACA	GTG	AAT	CCA	AAA	GAA	AGT	ATA	CTG	AAA	AGA	GTG	AAG	GAT	ATA	GGA	1104		
Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly			
				355					360				365					
TAC	ATC	TTT	AAA	GAG	AAA	TTT	GCT	AAA	GCT	GTG	GGA	CAG	GGT	TGT	GTC	1152		
Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys	Ala	Val	Gly	Gln	Gly	Cys	Val			
				370					375				380					
GAA	ATT	GGA	TCA	CAG	CGA	TAC	AAA	CTT	GGA	GTT	CGC	TTG	TAT	TAC	CGA	1200		
Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg			
				385					390				395					

GTA ATG GAA TCC ATG CTT AAA TCA GAA GAA GAA CGA TTA TCC ATT CAA	1248
Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln	
400 405 410	
AAT TTT AGC AAA CTT CTG AAT GAC AAC ATT TTT CAT ATG TCT TTA TTG	1296
Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu	
415 420 425 430	
GCG TGC GCT CTT GAG GTT GTA ATG GCC ACA TAT AGC AGA AGT ACA TCT	1344
Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser	
435 440 445	
CAG AAT CTT GAT TCT GGA ACA GAT TTG TCT TTC CCA TGG ATT CTG AAT	1392
Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn	
450 455 460	
GTG CTT AAT TTA AAA GCC TTT GAT TTT TAC AAA GTG ATC GAA AGT TTT	1440
Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe	
465 470 475	
ATC AAA GCA GAA GGC AAC TTG ACA AGA GAA ATG ATA AAA CAT TTA GAA	1488
Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu Glu	
480 485 490	
CGA TGT GAA CAT CGA ATC ATG GAA TCC CTT GCA TGG CTC TCA GAT TCA	1536
Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp Ser	
495 500 505 510	
CCT TTA TTT GAT CTT ATT AAA CAA TCA AAG GAC CGA GAA GGA CCA ACT	1584
Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr	
515 520 525	
GAT CAC CTT GAA TCT GCT TGT CCT CTT AAT CTT CCT CTC CAG AAT AAT	1632
Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn	
530 535 540	
CAC ACT GCA GCA GAT ATG TAT CTT TCT CCT GTA AGA TCT CCA AAG AAA	1680
His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys	
545 550 555	
AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT GCA AAT GCA GAG ACA CAA	1728
Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr Gln	
560 565 570	
GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA TTG AAA TCT ACC TCT CTT	1776
Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu	
575 580 585 590	
TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA GCC TAT CTC CGG CTA AAT	1824
Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn	
595 600 605	

ACA Thr	CTT Leu	TGT Cys	GAA Glu 610	CGC Arg	CTT Leu	CTG Leu	TCT Ser	GAG Glu 615	CAC His	CCA Pro	GAA Glu	TTA Leu	GAA Glu 620	CAT His	ATC Ile	1872
ATC Ile	TGG Trp	ACC Thr	CTT Leu 625	TTC Phe	CAG Gln	CAC His	ACC Thr 630	CTG Leu	CAG Gln	AAT Asn	GAG Glu	TAT Tyr 635	GAA Glu	CTC Leu	ATG Met	1920
AGA Arg	GAC Asp 640	AGG Arg	CAT His	TTG Leu	GAC Asp	CAA Gln 645	ATT Ile	ATG Met	ATG Met	TGT Cys	TCC Ser 650	ATG Met	TAT Tyr	GGC Gly	ATA Ile	1968
TGC Cys 655	AAA Lys	GTG Val	AAG Lys	AAT Asn 660	ATA Ile	GAC Asp	CTT Leu	AAA Lys	TTC Phe	AAA Lys 665	ATC Ile	ATT Ile	GTA Val	ACA Thr	GCA Ala 670	2016
TAC Tyr	AAG Lys	GAT Asp	CTT Leu 675	CCT Pro	CAT His	GCT Ala	GTT Val	CAG Gln 680	GAG Glu	ACA Thr	TTC Phe	AAA Lys	CGT Arg	GTT Val 685	TTG Leu	2064
ATC Ile	AAA Lys	GAA Glu 690	GAG Glu	GAG Glu	TAT Tyr	GAT Asp	TCT Ser 695	ATT Ile	ATA Ile	GTA Val	TTC Phe	TAT Tyr 700	AAC Asn	TCG Ser	GTC Val	2112
TTC Phe	ATG Met 705	CAG Gln	AGA Arg	CTG Leu	AAA Lys	ACA Thr	AAT Asn 710	ATT Ile	TTG Leu	CAG Gln	TAT Tyr 715	GCT Ala	TCC Ser	ACC Thr	AGG Arg	2160
CCC Pro	CCT Pro 720	ACC Thr	TTG Leu	TCA Ser	CCA Pro	ATA Ile 725	CCT Pro	CAC His	ATT Ile	CCT Pro	CGA Arg 730	AGC Ser	CCT Pro	TAC Tyr	AAG Lys	2208
TTT Phe 735	CCT Pro	AGT Ser	TCA Ser	CCC Pro	TTA Leu 740	CGG Arg	ATT Ile	CCT Pro	GGA Gly	GGG Gly 745	AAC Asn	ATC Ile	TAT Tyr	ATT Ile	TCA Ser 750	2256
CCC Pro	CTG Leu	AAG Lys	AGT Ser 755	CCA Pro	TAT Tyr	AAA Lys	ATT Ile	TCA Ser 760	GAA Glu	GGT Gly	CTG Leu	CCA Pro	ACA Thr	CCA Pro 765	ACA Thr	2304
AAA Lys	ATG Met	ACT Thr 770	CCA Pro	AGA Arg	TCA Ser	AGA Arg	ATC Ile 775	TTA Leu	GTA Val	TCA Ser	ATT Ile	GGT Gly 780	GAA Glu	TCA Ser	TTC Phe	2352
GGG Gly	ACT Thr 785	TCT Ser	GAG Glu	AAG Lys	TTC Phe 790	CAG Gln	AAA Lys	ATA Ile	AAT Asn	CAG Gln	ATG Met	GTA Val 795	TGT Cys	AAC Asn	AGC Ser	2400
GAC Asp	CGT Arg 800	GTG Val	CTC Leu	AAA Lys	AGA Arg	AGT Ser 805	GCT Ala	GAA Glu	GGA Gly	AGC Ser	AAC Asn 810	CCT Pro	CCT Pro	AAA Lys	CCA Pro	2448

CTG AAA AAA CTA CGC TTT GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA	2496
Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly	
815 820 825 830	
AGT AAA CAT CTC CCA GGA GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA	2544
Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu	
835 840 845	
ATG ACT TCT ACT CGA ACA CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC	2592
Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser	
850 855 860	
ATG GAT ACC TCA AAC AAG GAA GAG AAA TGAGGATCTC AGGACCTTGG	2639
Met Asp Thr Ser Asn Lys Glu Glu Lys	
865 870	
TGGACACTGT GTACACCTCT GGATTCATTG TCTCTCACAG ATGTGACTGT ATAACCTTCC	2699
CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAAATGTGC	2759
AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC	2819
AAGATTGAAA ATCTTGTGTA AATCCTGCCA TTTAAAAAGT TGTCAGCAGAT TGTTCCTCT	2879
TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG	2939
ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC	2999
TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAAATG	3059
TGTCCTATCT ATCTTCCAAA TGCAATTTGA TTGACTGCCC ATTCACCAAA ATTATCCTGA	3119
ACTCTTCTGC AAAAAATGGAT ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT	3179
AGATTTTATT TTAATTTATG AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT	3239
TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA TACTATCATA CTAATGAAAC	3299
AGATTTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT	3359
TTTTAAATGA GGATTATTGA TAGT	3383

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 871 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	1	5	10	15
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp	20	25	30	
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	35	40	45	
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	50	55	60	
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	65	70	75	80
Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	85	90	95	
Glu	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu	100	105	110	
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	115	120	125	
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	130	135	140	
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	145	150	155	160
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	165	170	175	
Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys	180	185	190	
Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	Met	195	200	205	
Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp	210	215	220	
Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys	225	230	235	240
Thr	Gly	Ser	Asn	Ser	Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	245	250	255	
Val	Glu	Asn	Leu	Ser	Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	260	265	270	
Asp	Leu	Asp	Ala	Arg	Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	275	280	285	

Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn
 290 295 300
 Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr Pro Val Arg Thr
 305 310 315 320
 Val Met Asn Thr Ile Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser
 325 330 335
 Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val
 340 345 350
 Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile
 355 360 365
 Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile
 370 375 380
 Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met
 385 390 395 400
 Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe
 405 410 415
 Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys
 420 425 430
 Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn
 435 440 445
 Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu
 450 455 460
 Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys
 465 470 475 480
 Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys
 485 490 495
 Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu
 500 505 510
 Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His
 515 520 525
 Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr
 530 535 540
 Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly
 545 550 555 560
 Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr
 565 570 575

Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu
 580 585 590
 Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu
 595 600 605
 Cys Glu Arg Leu Leu Ser Glu His Pro Glu Leu Glu His Ile Ile Trp
 610 615 620
 Thr Leu Phe Gln His Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp
 625 630 635 640
 Arg His Leu Asp Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys
 645 650 655
 Val Lys Asn Ile Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys
 660 665 670
 Asp Leu Pro His Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys
 675 680 685
 Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met
 690 695 700
 Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro
 705 710 715 720
 Thr Leu Ser Pro Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro
 725 730 735
 Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu
 740 745 750
 Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met
 755 760 765
 Thr Pro Arg Ser Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr
 770 775 780
 Ser Glu Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg
 785 790 795 800
 Val Leu Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys
 805 810 815
 Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys
 820 825 830
 His Leu Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr
 835 840 845
 Ser Thr Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp
 850 855 860

Thr Ser Asn Lys Glu Glu Lys
865 870

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3554 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2790

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GGT GAT ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	384
Gly Asp Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
115 120 125	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	432
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
130 135 140	

AAA	GTT	GAT	AAT	GCT	ATG	TCA	AGA	CTG	TTG	AAG	AAG	TAT	GAT	GTA	TTG	480
Lys	Val	Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	
		145					150					155				
TTT	GCA	CTC	TTC	AGC	AAA	TTG	GAA	AGG	ACA	TGT	GAA	CTT	ATA	TAT	TTG	528
Phe	Ala	Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	
	160					165				170						
ACA	CAA	CCC	AGC	AGT	TCG	ATA	TCT	ACT	GAA	ATA	AAT	TCT	GCA	TTG	GTG	576
Thr	Gln	Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	
	175				180				185					190		
CTA	AAA	GTT	TCT	TGG	ATC	ACA	TTT	TTA	TTA	GCT	AAA	GGG	GAA	GTA	TTA	624
Leu	Lys	Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	
			195					200						205		
CAA	ATG	GAA	GAT	GAT	CTG	GTG	ATT	TCA	TTT	CAG	TTA	ATG	CTA	TGT	GTC	672
Gln	Met	Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	
			210					215					220			
CTT	GAC	TAT	TTT	ATT	AAA	CTC	TCA	CCT	CCC	ATG	TTG	CTC	AAA	GAA	CCA	720
Leu	Asp	Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	
		225					230					235				
TAT	AAA	ACA	GCT	GTT	ATA	CCC	ATT	AAT	GGT	TCA	CCT	CGA	ACA	CCC	AGG	768
Tyr	Lys	Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	
	240					245				250						
CGA	GGT	CAG	AAC	AGG	AGT	GCA	CGG	ATA	GCA	AAA	CAA	CTA	GAA	AAT	GAT	816
Arg	Gly	Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	
	255				260				265					270		
ACA	AGA	ATT	ATT	GAA	GTT	CTC	TGT	AAA	GAA	CAT	GAA	TGT	AAT	ATA	GAT	864
Thr	Arg	Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	
				275				280						285		
GAG	GTG	AAA	AAT	GTT	TAT	TTC	AAA	AAT	TTT	ATA	CCT	TTT	ATG	AAT	TCT	912
Glu	Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	
			290					295					300			
CTT	GGA	CTT	GTA	ACA	TCT	AAT	GGA	CTT	CCA	GAG	GTT	GAA	AAT	CTT	TCT	960
Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	
		305					310					315				
AAA	CGA	TAC	GAA	GAA	ATT	TAT	CTT	AAA	AAT	AAA	GAT	CTA	GAT	GCA	AGA	1008
Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	
	320					325					330					
TTA	TTT	TTG	GAT	CAT	GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	ATA	GAC	AGT	1056
Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	
	335					340				345					350	

TTT	GAA	ACA	CAG	AGA	ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	GAA	GAG	GTG	1104	
Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val		
				355						360						365	
AAT	GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	AAC	ACT	ATC	1152	
Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile		
				370						375						380	
CAA	CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	CCT	TCA	GAA	1200	
Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu		
				385						390						395	
AAT	CTG	ATT	TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	AGT	1248	
Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser		
				400						405						410	
ATA	CTG	AAA	AGA	GTG	AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	GAG	AAA	TTT	1296	
Ile	Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe		
415				420				425				430					
GCT	AAA	GCT	GTG	GGA	CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	CAG	CGA	TAC	1344	
Ala	Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr		
				435						440						445	
AAA	CTT	GGA	GTT	CGC	TTG	TAT	TAC	CGA	GTA	ATG	GAA	TCC	ATG	CTT	AAA	1392	
Lys	Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys		
				450						455						460	
TCA	GAA	GAA	GAA	CGA	TTA	TCC	ATT	CAA	AAT	TTT	AGC	AAA	CTT	CTG	AAT	1440	
Ser	Glu	Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn		
				465						470						475	
GAC	AAC	ATT	TTT	CAT	ATG	TCT	TTA	TTG	GCG	TGC	GCT	CTT	GAG	GTT	GTA	1488	
Asp	Asn	Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val		
				480						485						490	
ATG	GCC	ACA	TAT	AGC	AGA	AGT	ACA	TCT	CAG	AAT	CTT	GAT	TCT	GGA	ACA	1536	
Met	Ala	Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr		
495				500				505				510					
GAT	TTG	TCT	TTC	CCA	TGG	ATT	CTG	AAT	GTG	CTT	AAT	TTA	AAA	GCC	TTT	1584	
Asp	Leu	Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe		
				515						520						525	
GAT	TTT	TAC	AAA	GTG	ATC	GAA	AGT	TTT	ATC	AAA	GCA	GAA	GGC	AAC	TTG	1632	
Asp	Phe	Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu		
				530						535						540	
ACA	AGA	GAA	ATG	ATA	AAA	CAT	TTA	GAA	CGA	TGT	GAA	CAT	CGA	ATC	ATG	1680	
Thr	Arg	Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met		
				545						550						555	

GAA Glu	TCC Ser	CTT Leu	GCA Ala	TGG Trp	CTC Leu	TCA Ser	GAT Asp	TCA Ser	CCT Pro	TTA Leu	TTT Phe	GAT Asp	CTT Leu	ATT Ile	AAA Lys	1728
560						565			570							
CAA Gln	TCA Ser	AAG Lys	GAC Asp	CGA Arg	GAA Glu	GGA Gly	CCA Pro	ACT Thr	GAT Asp	CAC His	CTT Leu	GAA Glu	TCT Ser	GCT Ala	TGT Cys	1776
575			580						585			590				
CCT Pro	CTT Leu	AAT Asn	CTT Leu	CCT Pro	CTC Leu	CAG Gln	AAT Asn	AAT Asn	CAC His	ACT Thr	GCA Ala	GCA Ala	GAT Asp	ATG Met	TAT Tyr	1824
			595						600			605				
CTT Leu	TCT Ser	CCT Pro	GTA Val	AGA Arg	TCT Ser	CCA Pro	AAG Lys	AAA Lys	AAA Lys	GGT Gly	TCA Ser	ACT Thr	ACG Thr	CGT Arg	GTA Val	1872
			610			615						620				
AAT Asn	TCT Ser	ACT Thr	GCA Ala	AAT Asn	GCA Ala	GAG Glu	ACA Thr	CAA Gln	GCA Ala	ACC Thr	TCA Ser	GCC Ala	TTC Phe	CAG Gln	ACC Thr	1920
625						630						635				
CAG Gln	AAG Lys	CCA Pro	TTG Leu	AAA Lys	TCT Ser	ACC Thr	TCT Ser	CTT Leu	TCA Ser	CTG Leu	TTT Phe	TAT Tyr	AAA Lys	AAA Lys	GTG Val	1968
640						645			650							
TAT Tyr	CGG Arg	CTA Leu	GCC Ala	TAT Tyr	CTC Leu	CGG Arg	CTA Leu	AAT Asn	ACA Thr	CTT Leu	TGT Cys	GAA Glu	CGC Arg	CTT Leu	CTG Leu	2016
655			660						665			670				
TCT Ser	GAG Glu	CAC His	CCA Pro	GAA Glu	TTA Leu	GAA Glu	CAT His	ATC Ile	ATC Ile	TGG Trp	ACC Thr	CTT Leu	TTC Phe	CAG Gln	CAC His	2064
			675						680			685				
ACC Thr	CTG Leu	CAG Gln	AAT Asn	GAG Glu	TAT Tyr	GAA Glu	CTC Leu	ATG Met	AGA Arg	GAC Asp	AGG Arg	CAT His	TTG Leu	GAC Asp	CAA Gln	2112
			690			695						700				
ATT Ile	ATG Met	ATG Met	TGT Cys	TCC Ser	ATG Met	TAT Tyr	GGC Gly	ATA Ile	TGC Cys	AAA Lys	GTG Val	AAG Lys	AAT Asn	ATA Ile	GAC Asp	2160
705						710						715				
CTT Leu	AAA Lys	TTC Phe	AAA Lys	ATC Ile	ATT Ile	GTA Val	ACA Thr	GCA Ala	TAC Tyr	AAG Lys	GAT Asp	CTT Leu	CCT Pro	CAT His	GCT Ala	2208
720						725			730							
GTT Val	CAG Gln	GAG Glu	ACA Thr	TTC Phe	AAA Lys	CGT Arg	GTT Val	TTG Leu	ATC Ile	AAA Lys	GAA Glu	GAG Glu	GAG Glu	TAT Tyr	GAT Asp	2256
735			740						745			750				
TCT Ser	ATT Ile	ATA Ile	GTA Val	TTC Phe	TAT Tyr	AAC Asn	TCG Ser	GTC Val	TTC Phe	ATG Met	CAG Gln	AGA Arg	CTG Leu	AAA Lys	ACA Thr	2304
			755						760			765				

AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA	2352
Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile	
770 775 780	
CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG	2400
Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg	
785 790 795	
ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA	2448
Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys	
800 805 810	
ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA	2496
Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg	
815 820 825 830	
ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG	2544
Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln	
835 840 845	
AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT	2592
Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser	
850 855 860	
GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT	2640
Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp	
865 870 875	
ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG	2688
Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu	
880 885 890	
TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA	2736
Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg	
895 900 905 910	
ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA	2784
Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu	
915 920 925	
GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG	2840
Glu Lys	
TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC	2900
TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC	2960
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA	3020
TTTAAAAAGT TGTAGCAGAT TGTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT	3080
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT	3140

TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT 3200
 TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCCTATCT ATCTTCCAAA TGCAATTTGA 3260
 TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAAATGGAT ATTATTAGAA 3320
 ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT 3380
 ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT 3440
 AACCATATGA TACTATCATA CTA CTACTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAAGT 3500
 TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT 3554

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 928 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	1	5	10	15
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp	20	25	30	
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	35	40	45	
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	50	55	60	
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	65	70	75	80
Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	85	90	95	
Glu	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Gly	Asp	100	105	110	
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	115	120	125	
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	130	135	140	

Phe	Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	
				725					730					735		
Glu	Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	
			740					745					750			
Ile	Val	Phe	Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	
		755					760					765				
Leu	Gln	Tyr	Ala	Ser	Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	
	770					775					780					
Ile	Pro	Arg	Ser	Pro	Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	
785					790				795						800	
Gly	Gly	Asn	Ile	Tyr	Ile	Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	
			805						810					815		
Glu	Gly	Leu	Pro	Thr	Pro	Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu	
		820						825					830			
Val	Ser	Ile	Gly	Glu	Ser	Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile	
		835					840					845				
Asn	Gln	Met	Val	Cys	Asn	Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu	
	850					855					860					
Gly	Ser	Asn	Pro	Pro	Lys	Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu	
865					870					875					880	
Gly	Ser	Asp	Glu	Ala	Asp	Gly	Ser	Lys	His	Leu	Pro	Gly	Glu	Ser	Lys	
			885						890					895		
Phe	Gln	Gln	Lys	Leu	Ala	Glu	Met	Thr	Ser	Thr	Arg	Thr	Arg	Met	Gln	
			900					905					910			
Lys	Gln	Lys	Met	Asn	Asp	Ser	Met	Asp	Thr	Ser	Asn	Lys	Glu	Glu	Lys	
		915					920					925				

WHAT IS CLAIMED IS:

1. A DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, said modified retinoblastoma tumor suppressor protein comprising an N-terminal modification.

2. The DNA segment of claim 1, wherein said gene encodes a modified retinoblastoma tumor suppressor protein comprising an N-terminal region that comprises a first sequence region from which at least one amino acid has been deleted.

3. The DNA segment of claim 2, wherein at least two amino acids have been deleted from said first sequence region.

4. The DNA segment of claim 3, wherein at least about 25 amino acids have been deleted from said first sequence region.

5. The DNA segment of claim 4, wherein at least about 100 amino acids have been deleted from said first sequence region.

6. The DNA segment of claim 5, wherein at least about 150 amino acids have been deleted from said first sequence region.

7. The DNA segment of claim 6, wherein at least about 300 amino acids have been deleted from said first sequence region.

8. The DNA segment of claim 2, wherein said first sequence region is located:

- a) between about amino acid 1 and about amino acid 50;
- b) between about amino acid 51 and about amino acid 100;
- c) between about amino acid 101 and about amino acid 150;
- d) between about amino acid 151 and about amino acid 200;
- e) between about amino acid 201 and about amino acid 250;
- f) between about amino acid 251 and about amino acid 300;
- g) between about amino acid 1 and about amino acid 100;
- h) between about amino acid 51 and about amino acid 150;
- i) between about amino acid 101 and about amino acid 200;
- j) between about amino acid 151 and about amino acid 250;
- k) between about amino acid 201 and about amino acid 300;
- l) between about amino acid 1 and about amino acid 150;
- m) between about amino acid 51 and about amino acid 200;

n) between about amino acid 101 and about amino acid 250;

o) between about amino acid 151 and about amino acid 300;

p) between about amino acid 1 and about amino acid 200;

q) between about amino acid 51 and about amino acid 250;

r) between about amino acid 101 and about amino acid 300;

s) between about amino acid 1 and about amino acid 250;

t) between about amino acid 51 and about amino acid 300; or

u) between about amino acid 1 and about amino acid 300.

9. The DNA segment of claim 2, wherein:

a) about amino acid 2 through about amino acid 34 have been deleted from said first sequence region;

b) about amino acid 2 through about amino acid 55 have been deleted from said first sequence region;

c) about amino acid 2 through about amino acid 78 have been deleted from said first sequence region;

d) about amino acid 2 through about amino acid 97 have been deleted from said first sequence region;

- e) about amino acid 2 through about amino acid 148 have been deleted from said first sequence region;
- 5 f) about amino acid 31 through about amino acid 107 have been deleted from said first sequence region;
- g) about amino acid 77 through about amino acid 107 have been deleted from said first sequence region;
- 10 h) about amino acid 111 through about amino acid 181 have been deleted from said first sequence region;
- i) about amino acid 111 through about amino acid 241 have been deleted from said first sequence region;
- 15 j) about amino acid 181 through about amino acid 241 have been deleted from said first sequence region; or
- 20 k) about amino acid 242 through about amino acid 300 have been deleted from said first sequence region.

10. The DNA segment of claim 2, wherein said N-terminal region of said modified
25 retinoblastoma tumor suppressor protein further comprises a second sequence region from which at least one amino acid has been deleted.

11. The DNA segment of claim 10, wherein about amino acid 2 through about amino acid 34,
30 and about amino acid 76 through about amino acid 112 have been deleted.

12. The DNA segment of claim 10, wherein about amino acid 2 through about amino acid 55, and about amino acid 76 through about amino acid 112 have been deleted.

5

13. The DNA segment of claim 1, wherein said gene encodes a modified retinoblastoma tumor suppressor protein comprising at least a first N-terminal mutation, and wherein said modified retinoblastoma tumor suppressor protein has an increased biological activity in comparison to the biological activity of the corresponding wild-type retinoblastoma tumor suppressor protein.

10

14. The DNA segment of claim 13, wherein said gene encodes a modified retinoblastoma tumor suppressor protein comprising a mutation at position 111.

15

15. The DNA segment of claim 14, wherein said modified retinoblastoma tumor suppressor protein comprises glycine at position 111 in place of aspartic acid.

20

16. The DNA segment of claim 13, wherein said modified retinoblastoma tumor suppressor protein comprises at least a second N-terminal mutation.

25

17. The DNA segment of claim 16, wherein said gene encodes a modified retinoblastoma tumor suppressor protein comprising a mutation at position 111 and a mutation at position 112.

18. The DNA segment of claim 17, wherein said modified retinoblastoma tumor suppressor protein comprises glycine at position 111 in place of aspartic acid, and aspartic acid at position 112 in place of glutamic acid.

19. The DNA segment of claim 1, wherein said gene encodes a modified retinoblastoma tumor suppressor protein comprising an N-terminal region from which at least one amino acid has been deleted, and which contains at least one amino acid mutation.

20. The DNA segment of claim 2, wherein said gene encodes a modified retinoblastoma tumor suppressor protein that comprises at least the C-terminal amino acid sequence from about position 370 to about position 928 of SEQ ID NO:2.

21. The DNA segment of claim 2, wherein said gene encodes a modified retinoblastoma tumor suppressor protein comprising the contiguous amino acid sequence of SEQ ID NO:29; SEQ ID NO:31; SEQ ID NO:33; SEQ ID NO:35; SEQ ID NO:37; SEQ ID NO:39; SEQ ID NO:41; SEQ ID NO:43; SEQ ID NO:45; SEQ ID NO:47; SEQ ID NO:49; or SEQ ID NO:51.

22. The DNA segment of claim 2, wherein said gene comprises the contiguous nucleic acid sequence from between position 7 and position 2691 of SEQ ID NO:28; from between position 7 and position 2628 of SEQ ID NO:30; from between position 7 and position 2559 of SEQ ID NO:32; from between position 7 and position 2502 of SEQ ID NO:34; from between position 7 and position 2349 of SEQ ID NO:36; from between position 7 and position 2559 of SEQ ID NO:38; from between position 7 and position 2697 of SEQ ID NO:40; from between position 7 and position 2583 of SEQ ID NO:42; from between position 7 and position 2397 of SEQ ID NO:44; from between position 7 and position 2613 of SEQ ID NO:46; from between position 7

and position 2619 of SEQ ID NO:48; or from between position 7 and position 2790 of SEQ ID NO:50.

5 23. The DNA segment of claim 1, operationally positioned under the control of a promoter.

24. The DNA segment of claim 23, further defined as a recombinant vector.

10

25. The DNA segment of claim 24, wherein said recombinant vector is comprised within an adenoviral vector.

15 26. The DNA segment of claim 25, wherein said adenoviral vector is comprised within a recombinant adenovirus.

20

27. The DNA segment of claim 1, comprised within a host cell.

28. The DNA segment of claim 27, wherein said host cell is a eukaryotic cell.

25

29. The DNA segment of claim 28, wherein said host cell is a human cell.

30. The DNA segment of claim 28, wherein said host cell is a tumor cell.

30

- 5
- 10
- 15
- 20
- 25
- 30
31. The DNA segment of claim 28, wherein said host cell is comprised within an animal.
32. The DNA segment of claim 31, wherein said animal is a human subject.
33. The DNA segment of claim 1, dispersed in a pharmaceutically acceptable excipient.
34. The DNA segment of claim 1, wherein said modified retinoblastoma tumor suppressor protein is characterized as:
- a) comprising an N-terminal region that comprises at least a first sequence region from which at least one amino acid has been deleted, and wherein said modified retinoblastoma tumor suppressor protein has a biological activity at least about equivalent to the biological activity of the corresponding wild-type retinoblastoma tumor suppressor protein; or
 - b) comprising an N-terminal region that comprises a first sequence region comprising at least one mutation, and wherein said modified retinoblastoma tumor suppressor protein has an increased biological activity in comparison to the biological activity of the corresponding wild-type retinoblastoma tumor suppressor protein.
35. A modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, said modified retinoblastoma tumor suppressor protein comprising an N-terminal modification, wherein said modified retinoblastoma tumor suppressor protein has a biological activity at least about equivalent to the biological activity of the corresponding wild-type retinoblastoma tumor suppressor protein.

36. A recombinant host cell comprising a DNA segment comprising an isolated gene encoding a modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, said modified retinoblastoma tumor suppressor protein comprising an N-terminal modification.

37. The recombinant host cell of claim 36, wherein said host cell is a tumor cell.

38. A method of inhibiting cellular proliferation, comprising contacting a cell with an effective inhibitory amount of a first modified retinoblastoma tumor suppressor protein other than pRB⁹⁴, said modified retinoblastoma tumor suppressor protein comprising an N-terminal modification.

39. The method of claim 38, wherein said cell is contacted with said first modified retinoblastoma tumor suppressor protein by providing to said cell a DNA segment that expresses said first modified retinoblastoma tumor suppressor protein in said cell.

40. The method of claim 38, wherein said cell is located within an animal and said first modified retinoblastoma tumor suppressor protein, or a gene encoding said modified retinoblastoma tumor suppressor protein, is administered to said animal in a pharmaceutically acceptable vehicle.

41. The method of claim 38, wherein said cell is contacted with a modified retinoblastoma tumor suppressor protein and a p53 tumor suppressor protein in a combined amount effective to inhibit cellular proliferation in said cell.

42. A method of inhibiting cellular proliferation, comprising contacting a cell with a retinoblastoma protein and a p53 protein in a combined amount effective to inhibit cellular proliferation in said cell.

43. A method of treating cancer, comprising administering to an animal with cancer a pharmaceutically acceptable composition comprising a biologically effective inhibitory amount of a first modified retinoblastoma tumor suppressor protein, other than pRB⁹⁴, that comprises an N-terminal modification.

ABSTRACT

Disclosed are modified broad-spectrum retinoblastoma tumor suppressor proteins that have at least the same, and in most cases higher biological activity than the corresponding wild-type retinoblastoma tumor suppressor protein. Exemplary modified retinoblastoma tumor suppressor proteins have a modified N-terminal region, in particular comprising one or more deletions and/or mutations. Also disclosed are methods of making and using the modified retinoblastoma tumor suppressor proteins, particularly in circumstances where inhibition of cell growth is desired. Thus the present disclosure provides methods for treating diseases, as exemplified by, but not limited to cancer, that are characterized by abnormal cellular proliferation.

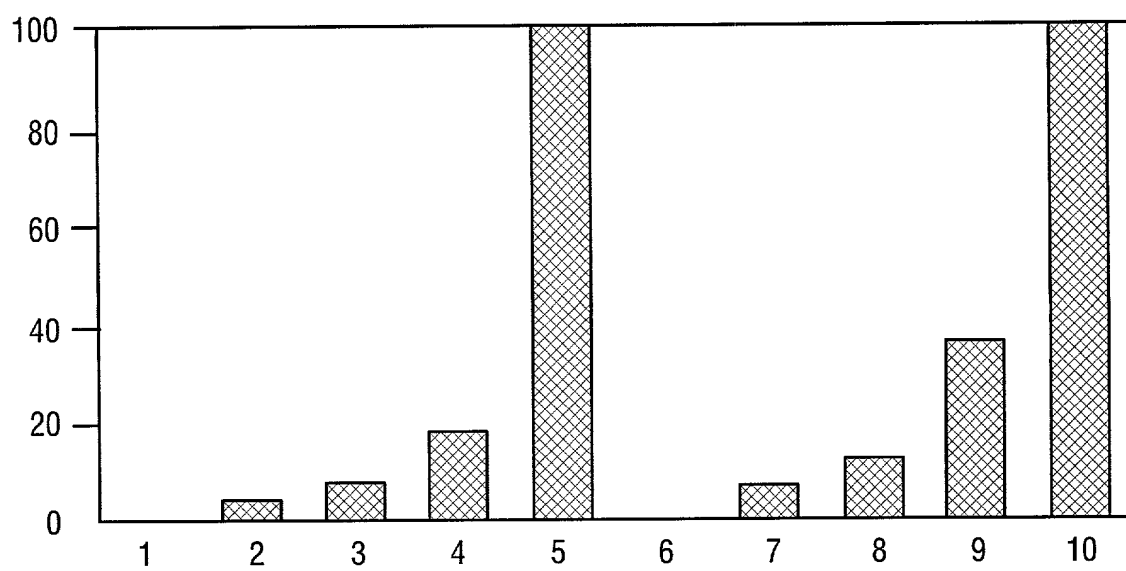


FIG. 1

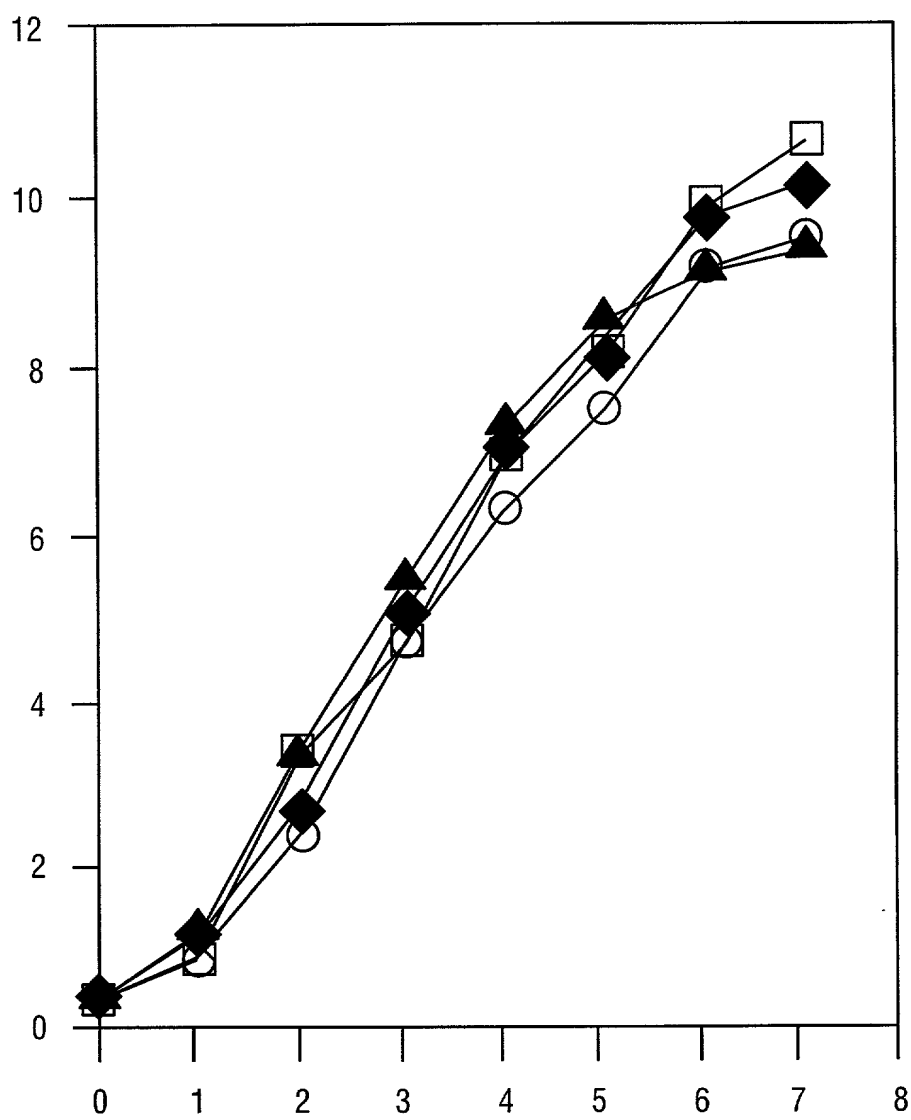


FIG. 2

FIG. 3A

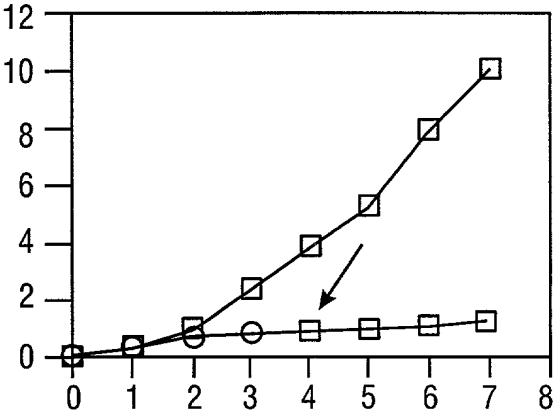


FIG. 3B

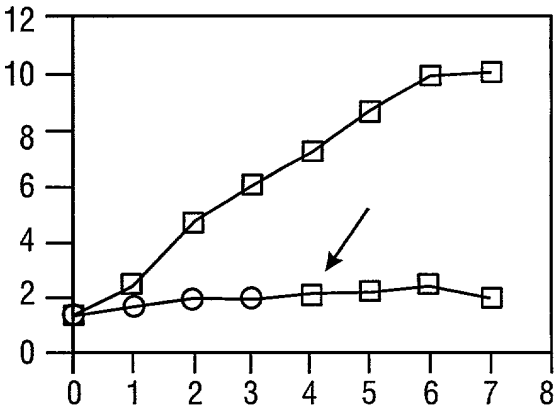


FIG. 3C

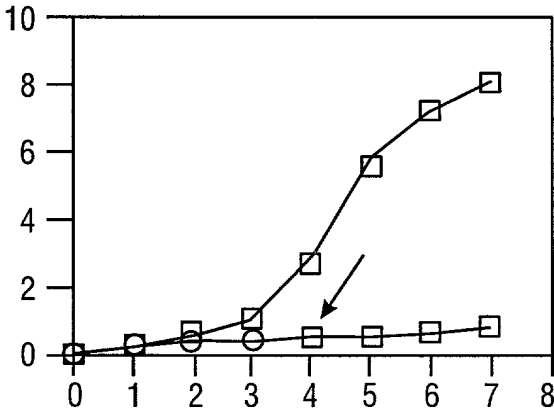


FIG. 4A

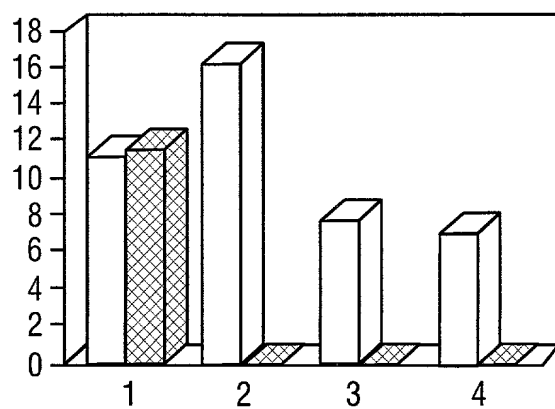


FIG. 4B

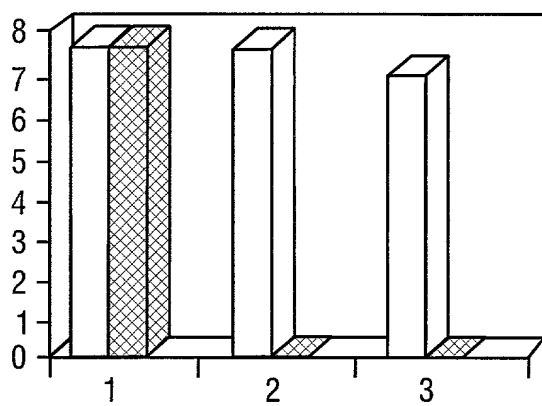
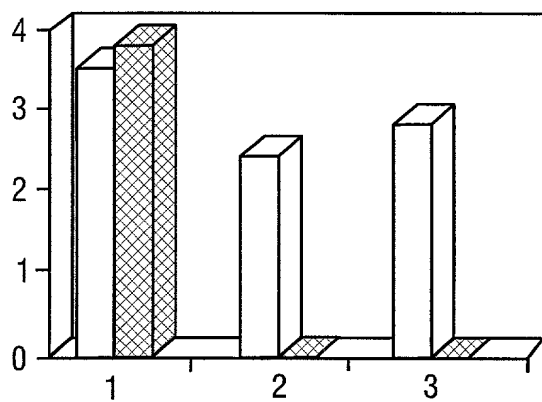


FIG. 4C



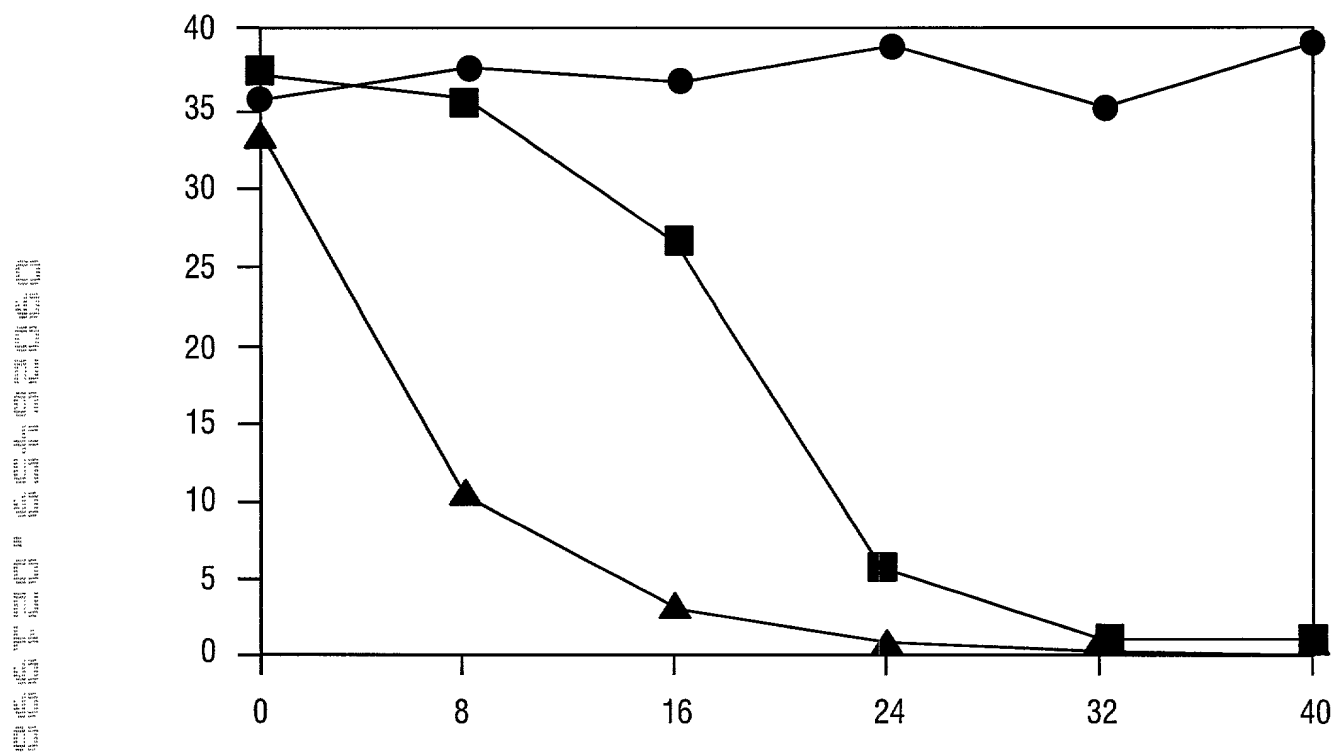


FIG. 5

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Hong-Ji Xu
Shi-Xue Hu
William F. Benedict
Yunli Zhou

Group Art Unit: Unknown

Examiner: Unknown

Atty. Dkt. No.: UTXC:506/HIB

Serial No.: Unknown

Filed: February 19, 1998

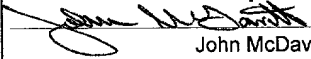
For: MODIFIED RETINOBLASTOMA
TUMOR SUPPRESSOR PROTEINS

EXPRESS MAIL MAILING LABEL

NUMBER EM 423 823 788 US

DATE OF DEPOSIT February 19, 1998

I hereby certify that this paper or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington D.C. 20231.


John McDavitt

STATEMENT AS REQUIRED UNDER 37 C.F.R. §§ 1.825(f) and (g)

ATTN: BOX SEQUENCE

Assistant Commissioner for Patents
Washington, DC 20231

Sir:

Submitted herewith is a computer readable form and a paper copy of the sequence listing of those sequences in the captioned patent application. The computer readable form of the sequence listing is the same as the paper copy of the sequence listing. The sequence information provided in the Specification is also the same as the sequence listing of the enclosed computer readable and paper forms of the sequence listing.

In accordance with 37 C.F.R. § 1.821(g), it is herewith represented that no new matter is included with this submission.

Respectfully submitted,

David W. Hibler

David W. Hibler
Reg. No. 41,071
Agent for Applicants

ARNOLD, WHITE & DURKEE
P.O. Box 4433
Houston, Texas 77210-4433
(512) 418-3000

Date: February 19, 1998

[illegible]

(1) GENERAL INFORMATION:

(i) APPLICANT: Xu, Hong-Ji
Hu, Shi-Xue
Benedict, William F.
Zhou, Yunli

(ii) TITLE OF INVENTION: MODIFIED RETINOBLASTOMA TUMOR SUPPRESSOR
PROTEINS

(iii) NUMBER OF SEQUENCES: 51

(iv) CORRESPONDENCE ADDRESS:
 (A) ADDRESSEE: Arnold, White & Durkee
 (B) STREET: P.O. Box 4433
 (C) CITY: Houston
 (D) STATE: TX
 (E) COUNTRY: USA
 (F) ZIP: 77210-4433

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
 (A) APPLICATION NUMBER: US UNKNOWN
 (B) FILING DATE:
 (C) CLASSIFICATION: UNKNOWN

(vii) PRIOR APPLICATION DATA:
 (A) APPLICATION NUMBER: US 60/038,118
 (B) FILING DATE: 20-FEB-1997

(viii) ATTORNEY/AGENT INFORMATION:
 (A) NAME: Hibler, David W.
 (B) REGISTRATION NUMBER: 41,071
 (C) REFERENCE/DOCKET NUMBER: UTXC:506

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 512/418-3000
(B) TELEFAX: 512/474-7577

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3555 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2790

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	384
Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
115 120 125	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	432
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
130 135 140	
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG	480
Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu	
145 150 155	

TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu 160 165 170	528
ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT GCA TTG GTG Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val 175 180 185 190	576
CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu 195 200 205	624
CAA ATG GAA GAT GAT CTG GTG ATT TCA TTT CAG TTA ATG CTA TGT GTC Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val 210 215 220	672
CTT GAC TAT TTT ATT AAA CTC TCA CCT CCC ATG TTG CTC AAA GAA CCA Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro 225 230 235	720
TAT AAA ACA GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg 240 245 250	768
CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp 255 260 265 270	816
ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp 275 280 285	864
GAG GTG AAA AAT GTT TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser 290 295 300	912
CTT GGA CTT GTA ACA TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser 305 310 315	960
AAA CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg 320 325 330	1008
TTA TTT TTG GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser 335 340 345 350	1056
TTT GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val 355 360 365	1104

AAT GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile 370 375 380	1152
CAA CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu 385 390 395	1200
AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser 400 405 410	1248
ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe 415 420 425 430	1296
GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr 435 440 445	1344
AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys 450 455 460	1392
TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn 465 470 475	1440
GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val 480 485 490	1488
ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr 495 500 505 510	1536
GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe 515 520 525	1584
GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu 530 535 540	1632
ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met 545 550 555	1680
GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys 560 565 570	1728

CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys 575 580 585 590	1776
CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr 595 600 605	1824
CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val 610 615 620	1872
AAT TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr 625 630 635	1920
CAG AAG CCA TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val 640 645 650	1968
TAT CGG CTA GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu 655 660 665 670	2016
TCT GAG CAC CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC Ser Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His 675 680 685	2064
ACC CTG CAG AAT GAG TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln 690 695 700	2112
ATT ATG ATG TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp 705 710 715	2160
CTT AAA TTC AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala 720 725 730	2208
GTT CAG GAG ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp 735 740 745 750	2256
TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr 755 760 765	2304
AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile 770 775 780	2352

CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg 785 790 795	2400
ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys 800 805 810	2448
ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg 815 820 825 830	2496
ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln 835 840 845	2544
AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser 850 855 860	2592
GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp 865 870 875	2640
ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu 880 885 890	2688
TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg 895 900 905 910	2736
ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu 915 920 925	2784
GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG Glu Lys	2840
TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC	2900
TTTCTGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC	2960
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGCTGTA AATCCTGCCA	3020
TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT	3080
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT	3140
TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT	3200
TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTTGA	3260

TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA	3320
ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT	3380
ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT	3440
AACCATATGA TACTATCATA CTACTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT	3500
TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGTC	3555

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 928 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	1	5	10	15
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp	20	25	30	
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	35	40	45	
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	50	55	60	
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	65	70	75	80
Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	85	90	95	
Glu	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu	100	105	110	
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	115	120	125	
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	130	135	140	
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	145	150	155	160
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	165	170	175	

Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn
465 470 475 480

Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala
485 490 495

Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu
500 505 510

Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe
515 520 525

Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg
530 535 540

Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser
545 550 555 560

Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser
565 570 575

Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu
580 585 590

Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser
595 600 605

Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser
610 615 620

Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys
625 630 635 640

Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg
645 650 655

Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu
660 665 670

His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu
675 680 685

Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met
690 695 700

Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys
705 710 715 720

Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln
725 730 735

Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile
740 745 750

Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile
755 760 765

Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His
770 775 780

Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro
785 790 795 800

Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser
805 810 815

Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu
820 825 830

Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile
835 840 845

Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu
850 855 860

Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu
865 870 875 880

Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys
885 890 895

Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln
900 905 910

Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
915 920 925

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3218 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2454

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCCGTC ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC
Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile
1 5 10

AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr 15 20 25 30	96
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu 35 40 45	144
TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu 50 55 60	192
ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT GCA TTG GTG Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val 65 70 75	240
CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu 80 85 90	288
CAA ATG GAA GAT GAT CTG GTG ATT TCA TTT CAG TTA ATG CTA TGT GTC Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val 95 100 105 110	336
CTT GAC TAT TTT ATT AAA CTC TCA CCT CCC ATG TTG CTC AAA GAA CCA Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro 115 120 125	384
TAT AAA ACA GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg 130 135 140	432
CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp 145 150 155	480
ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp 160 165 170	528
GAG GTG AAA AAT GTT TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser 175 180 185 190	576
CTT GGA CTT GTA ACA TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser 195 200 205	624
AAA CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg 210 215 220	672

TTA TTT TTG GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser 225 230 235	720
TTT GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val 240 245 250	768
AAT GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile 255 260 265 270	816
CAA CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu 275 280 285	864
AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser 290 295 300	912
ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe 305 310 315	960
GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr 320 325 330	1008
AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys 335 340 345 350	1056
TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn 355 360 365	1104
GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val 370 375 380	1152
ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr 385 390 395	1200
GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe 400 405 410	1248
GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu 415 420 425 430	1296

ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met 435 440 445	1344
GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys 450 455 460	1392
CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys 465 470 475	1440
CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr 480 485 490	1488
CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val 495 500 505 510	1536
AAT TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr 515 520 525	1584
CAG AAG CCA TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val 530 535 540	1632
TAT CGG CTA GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu 545 550 555	1680
TCT GAG CAC CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC Ser Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His 560 565 570	1728
ACC CTG CAG AAT GAG TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln 575 580 585 590	1776
ATT ATG ATG TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp 595 600 605	1824
CTT AAA TTC AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala 610 615 620	1872
GTT CAG GAG ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp 625 630 635	1920

TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr 640 645 650	1968
AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile 655 660 665 670	2016
CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg 675 680 685	2064
ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys 690 695 700	2112
ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg 705 710 715	2160
ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln 720 725 730	2208
AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser 735 740 745 750	2256
GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp 755 760 765	2304
ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu 770 775 780	2352
TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg 785 790 795	2400
ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu 800 805 810	2448
GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG Glu Lys 815	2504
TCTCTCACAG ATGTGACTGT ATAACTTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC	2564
TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC	2624
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA	2684

TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT	2744
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT	2804
TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT	2864
TTAATTTAAC ATGAACACCC TTAGAAAATG TGCCTATCT ATCTTCCAAA TGCAATTTGA	2924
TTGACTGCCC ATTACACAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA	2984
ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT	3044
ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT	3104
AACCATATGA TACTATCATA CTA CTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT	3164
TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT	3218

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 816 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile Ser Val	
1 5 10 15	
His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr Lys Val	
20 25 30	
Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala	
35 40 45	
Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln	
50 55 60	
Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys	
65 70 75 80	
Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met	
85 90 95	
Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val Leu Asp	
100 105 110	
Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys	
115 120 125	

6543210

Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly	130	135	140
Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	Arg	145	150	155
Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu	Val	165	170	175
Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu	Gly	180	185	190
Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	Lys	Arg	195	200	205
Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	Leu	Phe	210	215	220
Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	Phe	Glu	225	230	235
Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	Asn	Val	245	250	255
Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln	260	265	270
Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu	275	280	285
Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu	290	295	300
Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys	305	310	315
Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	325	330	335
Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu	340	345	350
Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	Asn	355	360	365
Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	370	375	380
Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	385	390	395
Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	405	410	415

Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg
 420 425 430
 Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser
 435 440 445
 Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser
 450 455 460
 Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu
 465 470 475 480
 Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser
 485 490 495
 Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser
 500 505 510
 Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys
 515 520 525
 Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg
 530 535 540
 Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu
 545 550 555 560
 His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu
 565 570 575
 Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met
 580 585 590
 Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys
 595 600 605
 Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln
 610 615 620
 Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile
 625 630 635 640
 Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile
 645 650 655
 Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His
 660 665 670
 Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro
 675 680 685
 Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser
 690 695 700

Glu	Gly	Leu	Pro	Thr	Pro	Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu
705					710					715					720
Val	Ser	Ile	Gly	Glu	Ser	Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile
			725						730					735	
Asn	Gln	Met	Val	Cys	Asn	Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu
			740					745					750		
Gly	Ser	Asn	Pro	Pro	Lys	Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu
		755					760					765			
Gly	Ser	Asp	Glu	Ala	Asp	Gly	Ser	Lys	His	Leu	Pro	Gly	Glu	Ser	Lys
		770				775					780				
Phe	Gln	Gln	Lys	Leu	Ala	Glu	Met	Thr	Ser	Thr	Arg	Thr	Arg	Met	Gln
785					790					795					800
Lys	Gln	Lys	Met	Asn	Asp	Ser	Met	Asp	Thr	Ser	Asn	Lys	Glu	Glu	Lys
				805					810					815	

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 285 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTCGAGCAAT GGGCGTGATA GCGGTTTGAC TCACGGGGAT TTCCAAGTCT CCACCCCAT	60
GACGTCAATG GGAGTTTGTT TTGGCACCAA AATCAACGGG ACTTTCCAAA ATGTCGTAAC	120
AACTCCGCCC CATTGACGCA AATGGGCGGT AGGCGTGAC GGTGGGAGGT CTATATAAGC	180
AGAGCTCGTT TAGTGAACCG TCAGATCGCC TGGAGACGCC ATCCACGCTG TTTTGACCTC	240
CATAGAAGAC ACCGGGACCG ATCCAGCCTC CGCGCCGCG AATTC	285

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGCTCGAGC AATGGGCGTG GATAGCGG

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCGCTCGAGC ACCAAAATCA ACGGGA

26

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCGCTCGAGC AACTCCGCCC CATTGAC

27

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TAGACATATG AATTCGCGGC C

21

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTAGAATTCG CTGTCTGCG

19

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCTCTAGATG CAGTTGGACC TGGGAG

26

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCCAAGCTTG CCGCCATGTC GTTCACTTTT AC

32

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTCCAAGAGA ATTCATAAAA GG

22

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCCAAGCTTG CCGCCATGGA GCAGGACAGC GGCCCGGAC

39

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCCAAGCTTG CCGCCATGGA TTTTACTGCA TTATGTCAG

39

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCCAAGCTTG CCGCCATGGA GAAAGTTTCA TCTTGTGAT

39

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCCAAGCTTG CCGCCATGCT GTGGGGAATC TGTATCTTT

39

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCCAAGCTTG CCGCCATGTC AAGACTGTTG AAGAAG

36

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCGCCTGAGG ACCTAGATGA GATGTCGTTC

30

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCGGTTAACC CTAGATGAGA TGTCGTTTAC T

31

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCCAAGCTTG CCGTCATGCC GCCCAAACC CCCC GA

36

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTCACCTAGG TCAACTGCTG CAAT

24

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTTGACCTAG GTGATATGTC GTTC

24

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCGCCTAGGA TCTACTGAAA TAAATTCTGC A

31

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCCGATATCA ACTGCTGGGT TGTGTCAAAT A

31

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCCGAATTCG TTTTATATGG TTCTTTGAGC AA

32

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
- (B) LOCATION: 4..5
- (D) OTHER INFORMATION: /note= "R=A or G"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCCRCCAUGG

10

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3455 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2691

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCCGTC ATG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG CTT	48
Met Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg Leu	
1 5 10	
GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG AAA	96
Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln Lys	
15 20 25 30	
TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG GAG	144
Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp Glu	
35 40 45	
AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG AAA	192
Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys Lys	
50 55 60	
AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA GAT	240
Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Asp	
65 70 75	

GAG Glu 80	ATG Met	TCG Ser	TTC Phe	ACT Thr	TTT Phe	ACT Thr	GAG Glu	CTA Leu	CAG Gln	AAA Lys	AAC Asn	ATA Ile	GAA Glu	ATC Ile	AGT Ser	288
GTC Val 95	CAT His	AAA Lys	TTC Phe	TTT Phe	AAC Asn	TTA Leu	CTA Leu	AAA Lys	GAA Glu	ATT Ile	GAT Asp	ACC Thr	AGT Ser	ACC Thr	AAA Lys	336
GTT Val	GAT Asp	AAT Asn	GCT Ala	ATG Met	TCA Ser	AGA Arg	CTG Leu	TTG Leu	AAG Lys	AAG Lys	TAT Tyr	GAT Asp	GTA Val	TTG Leu	TTT Phe	384
GCA Ala	CTC Leu	TTC Phe	AGC Ser	AAA Lys	TTG Leu	GAA Glu	AGG Arg	ACA Thr	TGT Cys	GAA Glu	CTT Leu	ATA Ile	TAT Tyr	TTG Leu	ACA Thr	432
CAA Gln	CCC Pro	AGC Ser	AGT Ser	TCG Ser	ATA Ile	TCT Ser	ACT Thr	GAA Glu	ATA Ile	AAT Asn	TCT Ser	GCA Ala	TTG Leu	GTG Val	CTA Leu	480
AAA Lys 160	GTT Val	TCT Ser	TGG Trp	ATC Ile	ACA Thr	TTT Phe	TTA Leu	TTA Leu	GCT Ala	AAA Lys	GGG Gly	GAA Glu	GTA Val	TTA Leu	CAA Gln	528
ATG Met 175	GAA Glu	GAT Asp	GAT Asp	CTG Leu	GTG Val	ATT Ile	TCA Ser	TTT Phe	CAG Gln	TTA Leu	ATG Met	CTA Leu	TGT Cys	GTC Val	CTT Leu	576
GAC Asp	TAT Tyr	TTT Phe	ATT Ile	AAA Lys	CTC Leu	TCA Ser	CCT Pro	CCC Pro	ATG Met	TTG Leu	CTC Leu	AAA Lys	GAA Glu	CCA Pro	TAT Tyr	624
AAA Lys 210	ACA Thr	GCT Ala	GTT Val	ATA Ile	CCC Pro	ATT Ile	AAT Asn	GGT Gly	TCA Ser	CCT Pro	CGA Arg	ACA Thr	CCC Pro	AGG Arg	CGA Arg	672
GGT Gly 225	CAG Gln	AAC Asn	AGG Arg	AGT Ser	GCA Ala	CGG Arg	ATA Ile	GCA Ala	AAA Lys	CAA Gln	CTA Leu	GAA Glu	AAT Asn	GAT Asp	ACA Thr	720
AGA Arg 240	ATT Ile	ATT Ile	GAA Glu	GTT Val	CTC Leu	TGT Cys	AAA Lys	GAA Glu	CAT His	GAA Glu	TGT Cys	AAT Asn	ATA Ile	GAT Asp	GAG Glu	768
GTG Val 255	AAA Lys	AAT Asn	GTT Val	TAT Tyr	TTC Phe	AAA Lys	AAT Asn	TTT Phe	ATA Ile	CCT Pro	TTT Phe	ATG Met	AAT Asn	TCT Ser	CTT Leu	816
GGA Gly 275	CTT Leu	GTA Val	ACA Thr	TCT Ser	AAT Asn	GGA Gly	CTT Leu	CCA Pro	GAG Glu	GTT Val	GAA Glu	AAT Asn	CTT Leu	TCT Ser	AAA Lys	864

CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA TTA	912
Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu	
290 295 300	
TTT TTG GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT TTT	960
Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe	
305 310 315	
GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG AAT	1008
Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn	
320 325 330	
GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC CAA	1056
Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln	
335 340 345 350	
CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA AAT	1104
Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn	
355 360 365	
CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT ATA	1152
Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile	
370 375 380	
CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT GCT	1200
Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala	
385 390 395	
AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC AAA	1248
Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys	
400 405 410	
CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA TCA	1296
Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser	
415 420 425 430	
GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT GAC	1344
Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp	
435 440 445	
AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA ATG	1392
Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met	
450 455 460	
GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA GAT	1440
Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp	
465 470 475	
TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT GAT	1488
Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp	
480 485 490	

TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG ACA Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr 495 500 505 510	1536
AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG GAA Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu 515 520 525	1584
TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA CAA Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln 530 535 540	1632
TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT CCT Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro 545 550 555	1680
CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT CTT Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu 560 565 570	1728
TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA AAT Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn 575 580 585 590	1776
TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC CAG Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln 595 600 605	1824
AAG CCA TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG TAT Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr 610 615 620	1872
CGG CTA GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG TCT Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser 625 630 635	1920
GAG CAC CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC ACC Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr 640 645 650	1968
CTG CAG AAT GAG TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA ATT Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile 655 660 665 670	2016
ATG ATG TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC CTT Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu 675 680 685	2064
AAA TTC AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT GTT Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val 690 695 700	2112

CAG Gln	GAG Glu	ACA Thr	TTC Phe	AAA Lys	CGT Arg	GTT Val	TTG Leu	ATC Ile	AAA Lys	GAA Glu	GAG Glu	GAG Glu	TAT Tyr	GAT Asp	TCT Ser	2160
705710715																
ATT Ile	ATA Ile	GTA Val	TTC Phe	TAT Tyr	AAC Asn	TCG Ser	GTC Val	TTC Phe	ATG Met	CAG Gln	AGA Arg	CTG Leu	AAA Lys	ACA Thr	AAT Asn	2208
720725730																
ATT Ile	TTG Leu	CAG Gln	TAT Tyr	GCT Ala	TCC Ser	ACC Thr	AGG Arg	CCC Pro	CCT Pro	ACC Thr	TTG Leu	TCA Ser	CCA Pro	ATA Ile	CCT Pro	2256
735740745750																
CAC His	ATT Ile	CCT Pro	CGA Arg	AGC Ser	CCT Pro	TAC Tyr	AAG Lys	TTT Phe	CCT Pro	AGT Ser	TCA Ser	CCC Pro	TTA Leu	CGG Arg	ATT Ile	2304
755760765																
CCT Pro	GGA Gly	GGG Gly	AAC Asn	ATC Ile	TAT Tyr	ATT Ile	TCA Ser	CCC Pro	CTG Leu	AAG Lys	AGT Ser	CCA Pro	TAT Tyr	AAA Lys	ATT Ile	2352
770775780																
TCA Ser	GAA Glu	GGT Gly	CTG Leu	CCA Pro	ACA Thr	CCA Pro	ACA Thr	AAA Lys	ATG Met	ACT Thr	CCA Pro	AGA Arg	TCA Ser	AGA Arg	ATC Ile	2400
785790795																
TTA Leu	GTA Val	TCA Ser	ATT Ile	GGT Gly	GAA Glu	TCA Ser	TTC Phe	GGG Gly	ACT Thr	TCT Ser	GAG Glu	AAG Lys	TTC Phe	CAG Gln	AAA Lys	2448
800805810																
ATA Ile	AAT Asn	CAG Gln	ATG Met	GTA Val	TGT Cys	AAC Asn	AGC Ser	GAC Asp	CGT Arg	GTG Val	CTC Leu	AAA Lys	AGA Arg	AGT Ser	GCT Ala	2496
815820825830																
GAA Glu	GGA Gly	AGC Ser	AAC Asn	CCT Pro	CCT Pro	AAA Lys	CCA Pro	CTG Leu	AAA Lys	AAA Lys	CTA Leu	CGC Arg	TTT Phe	GAT Asp	ATT Ile	2544
835840845																
GAA Glu	GGA Gly	TCA Ser	GAT Asp	GAA Glu	GCA Ala	GAT Asp	GGA Gly	AGT Ser	AAA Lys	CAT His	CTC Leu	CCA Pro	GGA Gly	GAG Glu	TCC Ser	2592
850855860																
AAA Lys	TTT Phe	CAG Gln	CAG Gln	AAA Lys	CTG Leu	GCA Ala	GAA Glu	ATG Met	ACT Thr	TCT Ser	ACT Thr	CGA Arg	ACA Thr	CGA Arg	ATG Met	2640
865870875																
CAA Gln	AAG Lys	CAG Gln	AAA Lys	ATG Met	AAT Asn	GAT Asp	AGC Ser	ATG Met	GAT Asp	ACC Thr	TCA Ser	AAC Asn	AAG Lys	GAA Glu	GAG Glu	2688
880885890																
AAA Lys 895																
TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG																
2741																
TCTCTCACAG ATGTGACTGT ATAACTTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC																
2801																

TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC	2861
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA	2921
TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT	2981
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT	3041
TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT	3101
TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTTGA	3161
TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA	3221
ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT	3281
ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT	3341
AACCATATGA TACTATCATA CTA CTGAAAC AGATTTTATA CCTCAGAATG TAAAAGAACT	3401
TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT	3455

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 895 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	Phe
1				5					10					15	
Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	Lys
			20					25					30		
Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	Val
		35					40					45			
Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	Glu
	50					55					60				
Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu	Met
65				70					75					80	
Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	His
			85					90						95	
Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	Asp
			100				105						110		

Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	Glu	
690						695					700					
Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	Ile	
705					710					715					720	
Val	Phe	Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	Leu	
				725					730					735		
Gln	Tyr	Ala	Ser	Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	Ile	
		740						745					750			
Pro	Arg	Ser	Pro	Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	Gly	
		755					760					765				
Gly	Asn	Ile	Tyr	Ile	Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	Glu	
	770					775					780					
Gly	Leu	Pro	Thr	Pro	Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu	Val	
785					790					795					800	
Ser	Ile	Gly	Glu	Ser	Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile	Asn	
				805					810					815		
Gln	Met	Val	Cys	Asn	Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu	Gly	
			820					825					830			
Ser	Asn	Pro	Pro	Lys	Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu	Gly	
		835					840					845				
Ser	Asp	Glu	Ala	Asp	Gly	Ser	Lys	His	Leu	Pro	Gly	Glu	Ser	Lys	Phe	
	850					855					860					
Gln	Gln	Lys	Leu	Ala	Glu	Met	Thr	Ser	Thr	Arg	Thr	Arg	Met	Gln	Lys	
865					870					875					880	
Gln	Lys	Met	Asn	Asp	Ser	Met	Asp	Thr	Ser	Asn	Lys	Glu	Glu	Lys		
			885						890					895		

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3392 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2628

Variable	Mean	SD	Min	Max
Age	34.5	10.2	21	55
Gender	Male	100%		
Education	High school	100%		
Occupation	Unemployed	100%		
Marital status	Married	100%		
Religion	Islam	100%		
Health status	Good	100%		
Smoking status	Non-smoker	100%		
Alcohol consumption	None	100%		
Family size	3.5	1.2	2	6
Income	Low	100%		
Health insurance	Yes	100%		
Access to healthcare	Yes	100%		
Healthcare utilization	Low	100%		
Healthcare satisfaction	Low	100%		
Healthcare access barriers	High	100%		
Healthcare quality	Low	100%		
Healthcare equity	Low	100%		
Healthcare sustainability	Low	100%		
Healthcare financing	Low	100%		
Healthcare management	Low	100%		
Healthcare information	Low	100%		
Healthcare communication	Low	100%		
Healthcare participation	Low	100%		
Healthcare accountability	Low	100%		
Healthcare transparency	Low	100%		
Healthcare integrity	Low	100%		
Healthcare ethics	Low	100%		
Healthcare law	Low	100%		
Healthcare policy	Low	100%		
Healthcare regulation	Low	100%		
Healthcare supervision	Low	100%		
Healthcare evaluation	Low	100%		
Healthcare research	Low	100%		
Healthcare innovation	Low	100%		
Healthcare development	Low	100%		
Healthcare reform	Low	100%		
Healthcare change	Low	100%		
Healthcare improvement	Low	100%		
Healthcare optimization	Low	100%		
Healthcare enhancement	Low	100%		
Healthcare advancement	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		
Healthcare expansion	Low	100%		
Healthcare development	Low	100%		
Healthcare progress	Low	100%		
Healthcare growth	Low	100%		

GCCATC	ATG	GAT	TTT	ACT	GCA	TTA	TGT	CAG	AAA	TTA	AAG	ATA	CCA	GAT		48
Met	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	Lys	Ile	Pro	Asp			
1				5					10							
CAT	GTC	AGA	GAG	AGA	GCT	TGG	TTA	ACT	TGG	GAG	AAA	GTT	TCA	TCT	GTG	96
His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	Val	Ser	Ser	Val	
15				20					25						30	
GAT	GGA	GTA	TTG	GGA	GGT	TAT	ATT	CAA	AAG	AAA	AAG	GAA	CTG	TGG	GGA	144
Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	Glu	Leu	Trp	Gly	
				35				40						45		
ATC	TGT	ATC	TTT	ATT	GCA	GCA	GTT	GAC	CTA	GAT	GAG	ATG	TCG	TTC	ACT	192
Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu	Met	Ser	Phe	Thr	
			50					55					60			
TTT	ACT	GAG	CTA	CAG	AAA	AAC	ATA	GAA	ATC	AGT	GTC	CAT	AAA	TTC	TTT	240
Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	His	Lys	Phe	Phe	
		65					70					75				
AAC	TTA	CTA	AAA	GAA	ATT	GAT	ACC	AGT	ACC	AAA	GTT	GAT	AAT	GCT	ATG	288
Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	Asp	Asn	Ala	Met	
	80					85					90					
TCA	AGA	CTG	TTG	AAG	AAG	TAT	GAT	GTA	TTG	TTT	GCA	CTC	TTC	AGC	AAA	336
Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	Leu	Phe	Ser	Lys	
95				100					105						110	
TTG	GAA	AGG	ACA	TGT	GAA	CTT	ATA	TAT	TTG	ACA	CAA	CCC	AGC	AGT	TCG	384
Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	Pro	Ser	Ser	Ser	
				115					120					125		
ATA	TCT	ACT	GAA	ATA	AAT	TCT	GCA	TTG	GTG	CTA	AAA	GTT	TCT	TGG	ATC	432
Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys	Val	Ser	Trp	Ile	
			130					135						140		
ACA	TTT	TTA	TTA	GCT	AAA	GGG	GAA	GTA	TTA	CAA	ATG	GAA	GAT	GAT	CTG	480
Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	Met	Glu	Asp	Asp	Leu	
		145					150					155				
GTG	ATT	TCA	TTT	CAG	TTA	ATG	CTA	TGT	GTC	CTT	GAC	TAT	TTT	ATT	AAA	528
Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp	Tyr	Phe	Ile	Lys	
	160					165					170					
CTC	TCA	CCT	CCC	ATG	TTG	CTC	AAA	GAA	CCA	TAT	AAA	ACA	GCT	GTT	ATA	576
Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys	Thr	Ala	Val	Ile	
175					180					185					190	
CCC	ATT	AAT	GGT	TCA	CCT	CGA	ACA	CCC	AGG	CGA	GGT	CAG	AAC	AGG	AGT	624
Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly	Gln	Asn	Arg	Ser	
				195					200					205		

GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT ACA AGA ATT ATT GAA GTT Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu Val 210 215 220	672
CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT GAG GTG AAA AAT GTT TAT Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val Tyr 225 230 235	720
TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT CTT GGA CTT GTA ACA TCT Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr Ser 240 245 250	768
AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT AAA CGA TAC GAA GAA ATT Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile 255 260 265 270	816
TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA TTA TTT TTG GAT CAT GAT Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp 275 280 285	864
AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT TTT GAA ACA CAG AGA ACA Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr 290 295 300	912
CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG AAT GTA ATT CCT CCA CAC Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro His 305 310 315	960
ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC CAA CAA TTA ATG ATG ATT Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met Ile 320 325 330	1008
TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA AAT CTG ATT TCC TAT TTT Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe 335 340 345 350	1056
AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT ATA CTG AAA AGA GTG AAG Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys 355 360 365	1104
GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT GCT AAA GCT GTG GGA CAG Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln 370 375 380	1152
GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC AAA CTT GGA GTT CGC TTG Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu 385 390 395	1200
TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA TCA GAA GAA GAA CGA TTA Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu 400 405 410	1248

TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT GAC AAC ATT TTT CAT ATG	1296
Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met	
415 420 425 430	
TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA ATG GCC ACA TAT AGC AGA	1344
Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg	
435 440 445	
AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA GAT TTG TCT TTC CCA TGG	1392
Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp	
450 455 460	
ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT GAT TTT TAC AAA GTG ATC	1440
Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile	
465 470 475	
GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG ACA AGA GAA ATG ATA AAA	1488
Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys	
480 485 490	
CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG GAA TCC CTT GCA TGG CTC	1536
His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu	
495 500 505 510	
TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA CAA TCA AAG GAC CGA GAA	1584
Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu	
515 520 525	
GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT CCT CTT AAT CTT CCT CTC	1632
Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu	
530 535 540	
CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT CTT TCT CCT GTA AGA TCT	1680
Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser	
545 550 555	
CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT GCA AAT GCA	1728
Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala	
560 565 570	
GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA TTG AAA TCT	1776
Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys Ser	
575 580 585 590	
ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA GCC TAT CTC	1824
Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr Leu	
595 600 605	
CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG TCT GAG CAC CCA GAA TTA	1872
Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His Pro Glu Leu	
610 615 620	

GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC ACC CTG CAG AAT GAG TAT Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln Asn Glu Tyr 625 630 635	1920
GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA ATT ATG ATG TGT TCC ATG Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met Met Cys Ser Met 640 645 650	1968
TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC CTT AAA TTC AAA ATC ATT Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys Phe Lys Ile Ile 655 660 665 670	2016
GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT GTT CAG GAG ACA TTC AAA Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln Glu Thr Phe Lys 675 680 685	2064
CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT TCT ATT ATA GTA TTC TAT Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe Tyr 690 695 700	2112
AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA AAT ATT TTG CAG TAT GCT Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr Ala 705 710 715	2160
TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA CCT CAC ATT CCT CGA AGC Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His Ile Pro Arg Ser 720 725 730	2208
CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG ATT CCT GGA GGG AAC ATC Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn Ile 735 740 745 750	2256
TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA ATT TCA GAA GGT CTG CCA Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu Pro 755 760 765	2304
ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA ATC TTA GTA TCA ATT GGT Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val Ser Ile Gly 770 775 780	2352
GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG AAA ATA AAT CAG ATG GTA Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn Gln Met Val 785 790 795	2400
TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT GCT GAA GGA AGC AAC CCT Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly Ser Asn Pro 800 805 810	2448
CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT ATT GAA GGA TCA GAT GAA Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp Glu 815 820 825 830	2496

GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG TCC AAA TTT CAG CAG AAA	2544
Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln Lys	
835 840 845	
CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA ATG CAA AAG CAG AAA ATG	2592
Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys Met	
850 855 860	
AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG AAA TGAGGATCTC	2638
Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys	
865 870	
AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG TCTCTCACAG ATGTGACTGT	2698
ATAACTTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT TTTGTGGATA	2758
TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG TTAGTCATTG	2818
TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA TTTAAAAAGT TGTAGCAGAT	2878
TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG GCCCTAGAGT	2938
GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG CATATAGGTG	2998
ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC ATGAACACCC	3058
TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTTGA TTGACTGCCC ATTCACCAAAA	3118
ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA AATTACTAAT	3178
TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT ACTGTGTGCT TGTTTTATAA	3238
AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA TACTATCATA	3298
CTACTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT TTTCTTCATC	3358
CAACTTATGT TTTTAAATGA GGATTATTGA TAGT	3392

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 874 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Asp Phe Thr Ala Leu Cys Gln Lys Leu Lys Ile Pro Asp His Val
1 5 10 15

Arg Glu Arg Ala Trp Leu Thr Trp Glu Lys Val Ser Ser Val Asp Gly
 20 25 30
 Val Leu Gly Gly Tyr Ile Gln Lys Lys Lys Glu Leu Trp Gly Ile Cys
 35 40 45
 Ile Phe Ile Ala Ala Val Asp Leu Asp Glu Met Ser Phe Thr Phe Thr
 50 55 60
 Glu Leu Gln Lys Asn Ile Glu Ile Ser Val His Lys Phe Phe Asn Leu
 65 70 75 80
 Leu Lys Glu Ile Asp Thr Ser Thr Lys Val Asp Asn Ala Met Ser Arg
 85 90 95
 Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala Leu Phe Ser Lys Leu Glu
 100 105 110
 Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln Pro Ser Ser Ser Ile Ser
 115 120 125
 Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val Ser Trp Ile Thr Phe
 130 135 140
 Leu Leu Ala Lys Gly Glu Val Leu Gln Met Glu Asp Asp Leu Val Ile
 145 150 155 160
 Ser Phe Gln Leu Met Leu Cys Val Leu Asp Tyr Phe Ile Lys Leu Ser
 165 170 175
 Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys Thr Ala Val Ile Pro Ile
 180 185 190
 Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg Ser Ala Arg
 195 200 205
 Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu Val Leu Cys
 210 215 220
 Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val Tyr Phe Lys
 225 230 235 240
 Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr Ser Asn Gly
 245 250 255
 Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile Tyr Leu
 260 265 270
 Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp Lys Thr
 275 280 285
 Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro Arg
 290 295 300

Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr Pro
 305 310 315 320

Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met Ile Leu Asn
 325 330 335

Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn
 340 345 350

Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile
 355 360 365

Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys
 370 375 380

Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr
 385 390 395 400

Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile
 405 410 415

Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu
 420 425 430

Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr
 435 440 445

Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu
 450 455 460

Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser
 465 470 475 480

Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu
 485 490 495

Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp
 500 505 510

Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro
 515 520 525

Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn
 530 535 540

Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys
 545 550 555 560

Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr
 565 570 575

Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys Ser Thr Ser
 580 585 590

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3323 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2559

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCCATC ATG GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT	48
Met Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr	
1 5 10	
ATT CAA AAG AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA	96
Ile Gln Lys Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala	
15 20 25 30	
GTT GAC CTA GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC	144
Val Asp Leu Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn	
35 40 45	
ATA GAA ATC AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT	192
Ile Glu Ile Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp	
50 55 60	
ACC AGT ACC AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT	240
Thr Ser Thr Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr	
65 70 75	
GAT GTA TTG TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT	288
Asp Val Leu Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu	
80 85 90	
ATA TAT TTG ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT	336
Ile Tyr Leu Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser	
95 100 105 110	
GCA TTG GTG CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG	384
Ala Leu Val Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly	
115 120 125	
GAA GTA TTA CAA ATG GAA GAT GAT CTG GTG ATT TCA TTT CAG TTA ATG	432
Glu Val Leu Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met	
130 135 140	
CTA TGT GTC CTT GAC TAT TTT ATT AAA CTC TCA CCT CCC ATG TTG CTC	480
Leu Cys Val Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu	
145 150 155	

AAA	GAA	CCA	TAT	AAA	ACA	GCT	GTT	ATA	CCC	ATT	AAT	GGT	TCA	CCT	CGA	528
Lys	Glu	Pro	Tyr	Lys	Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	
160						165					170					
ACA	CCC	AGG	CGA	GGT	CAG	AAC	AGG	AGT	GCA	CGG	ATA	GCA	AAA	CAA	CTA	576
Thr	Pro	Arg	Arg	Gly	Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	
175					180					185					190	
GAA	AAT	GAT	ACA	AGA	ATT	ATT	GAA	GTT	CTC	TGT	AAA	GAA	CAT	GAA	TGT	624
Glu	Asn	Asp	Thr	Arg	Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	
				195					200					205		
AAT	ATA	GAT	GAG	GTG	AAA	AAT	GTT	TAT	TTC	AAA	AAT	TTT	ATA	CCT	TTT	672
Asn	Ile	Asp	Glu	Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	
			210					215					220			
ATG	AAT	TCT	CTT	GGA	CTT	GTA	ACA	TCT	AAT	GGA	CTT	CCA	GAG	GTT	GAA	720
Met	Asn	Ser	Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	
		225					230					235				
AAT	CTT	TCT	AAA	CGA	TAC	GAA	GAA	ATT	TAT	CTT	AAA	AAT	AAA	GAT	CTA	768
Asn	Leu	Ser	Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	
	240						245				250					
GAT	GCA	AGA	TTA	TTT	TTG	GAT	CAT	GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	816
Asp	Ala	Arg	Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	
255					260					265					270	
ATA	GAC	AGT	TTT	GAA	ACA	CAG	AGA	ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	864
Ile	Asp	Ser	Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	
				275					280					285		
GAA	GAG	GTG	AAT	GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	912
Glu	Glu	Val	Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	
			290					295					300			
AAC	ACT	ATC	CAA	CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	960
Asn	Thr	Ile	Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	
		305					310					315				
CCT	TCA	GAA	AAT	CTG	ATT	TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	1008
Pro	Ser	Glu	Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	
		320				325					330					
AAA	GAA	AGT	ATA	CTG	AAA	AGA	GTG	AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	1056
Lys	Glu	Ser	Ile	Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	
335					340					345				350		
GAG	AAA	TTT	GCT	AAA	GCT	GTG	GGA	CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	1104
Glu	Lys	Phe	Ala	Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	
				355					360					365		

CAG CGA TAC AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser 370 375 380	1152
ATG CTT AAA TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys 385 390 395	1200
CTT CTG AAT GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu 400 405 410	1248
GAG GTT GTA ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp 415 420 425 430	1296
TCT GGA ACA GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu 435 440 445	1344
AAA GCC TTT GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu 450 455 460	1392
GGC AAC TTG ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His 465 470 475	1440
CGA ATC ATG GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp 480 485 490	1488
CTT ATT AAA CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu 495 500 505 510	1536
TCT GCT TGT CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala 515 520 525	1584
GAT ATG TAT CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr 530 535 540	1632
ACG CGT GTA AAT TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala 545 550 555	1680
TTC CAG ACC CAG AAG CCA TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT Phe Gln Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr 560 565 570	1728

AAA Lys 575	AAA Lys	GTG Val	TAT Tyr	CGG Arg	CTA Leu 580	GCC Ala	TAT Tyr	CTC Leu	CGG Arg 585	CTA Leu	AAT Asn	ACA Thr	CTT Leu	TGT Cys	GAA Glu 590	1776
CGC Arg	CTT Leu	CTG Leu	TCT Ser	GAG Glu 595	CAC His	CCA Pro	GAA Glu	TTA Leu 600	GAA Glu	CAT His	ATC Ile	ATC Ile	TGG Trp	ACC Thr 605	CTT Leu	1824
TTC Phe	CAG Gln	CAC His	ACC Thr 610	CTG Leu	CAG Gln	AAT Asn	GAG Glu 615	TAT Tyr 615	GAA Glu	CTC Leu	ATG Met	AGA Arg	GAC Asp 620	AGG Arg	CAT His	1872
TTG Leu	GAC Asp	CAA Gln 625	ATT Ile	ATG Met	ATG Met	TGT Cys 630	TCC Ser 630	ATG Met	TAT Tyr	GGC Gly	ATA Ile	TGC Cys 635	AAA Lys	GTG Val	AAG Lys	1920
AAT Asn 640	ATA Ile	GAC Asp	CTT Leu	AAA Lys	TTC Phe 645	AAA Lys 645	ATC Ile	ATT Ile	GTA Val	ACA Thr 650	GCA Ala 650	TAC Tyr	AAG Lys	GAT Asp	CTT Leu	1968
CCT Pro 655	CAT His	GCT Ala	GTT Val	CAG Gln 660	GAG Glu 660	ACA Thr	TTC Phe	AAA Lys	CGT Arg 665	GTT Val 665	TTG Leu	ATC Ile	AAA Lys	GAA Glu 670	GAG Glu	2016
GAG Glu	TAT Tyr	GAT Asp	TCT Ser	ATT Ile 675	ATA Ile	GTA Val	TTC Phe	TAT Tyr 680	AAC Asn 680	TCG Ser	GTC Val	TTC Phe	ATG Met	CAG Gln 685	AGA Arg	2064
CTG Leu	AAA Lys	ACA Thr	AAT Asn 690	ATT Ile	TTG Leu	CAG Gln	TAT Tyr 695	GCT Ala 695	TCC Ser	ACC Thr	AGG Arg	CCC Pro 700	CCT Pro 700	ACC Thr	TTG Leu	2112
TCA Ser 705	CCA Pro	ATA Ile 705	CCT Pro	CAC His	ATT Ile	CCT Pro	CGA Arg 710	AGC Ser 710	CCT Pro	TAC Tyr	AAG Lys	TTT Phe 715	CCT Pro	AGT Ser	TCA Ser	2160
CCC Pro 720	TTA Leu	CGG Arg	ATT Ile	CCT Pro	GGA Gly 725	GGG Gly 725	AAC Asn	ATC Ile	TAT Tyr	ATT Ile 730	TCA Ser 730	CCC Pro	CTG Leu	AAG Lys	AGT Ser	2208
CCA Pro 735	TAT Tyr	AAA Lys	ATT Ile	TCA Ser 740	GAA Glu 740	GGT Gly	CTG Leu	CCA Pro	ACA Thr 745	CCA Pro 745	ACA Thr	AAA Lys	ATG Met	ACT Thr 750	CCA Pro 750	2256
AGA Arg	TCA Ser	AGA Arg	ATC Ile 755	TTA Leu 755	GTA Val	TCA Ser	ATT Ile 760	GGT Gly 760	GAA Glu 760	TCA Ser 760	TTC Phe	GGG Gly	ACT Thr 765	TCT Ser 765	GAG Glu 765	2304
AAG Lys	TTC Phe	CAG Gln 770	AAA Lys	ATA Ile	AAT Asn	CAG Gln	ATG Met 775	GTA Val 775	TGT Cys 775	AAC Asn 775	AGC Ser	GAC Asp 780	CGT Arg 780	GTG Val	CTC Leu	2352

AAA AGA AGT GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA	2400
Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu	
785 790 795	
CGC TTT GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC	2448
Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu	
800 805 810	
CCA GGA GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT	2496
Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr	
815 820 825 830	
CGA ACA CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA	2544
Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser	
835 840 845	
AAC AAG GAA GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT	2599
Asn Lys Glu Glu Lys	
850	
GGATTTCATTG TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC	2659
ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAATGTGC AGATGCAATT GTTTGGGTGA	2719
TTCCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA	2779
AATCCTGCCA TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG	2839
CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC	2899
TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT	2959
GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA	3019
TGCAATTTGA TTGACTGCCC ATTCACCAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT	3079
ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG	3139
AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA	3199
AGCAAAGTAT AACCATATGA TACTATCATA CTACTGAAAC AGATTCATA CCTCAGAATG	3259
TAAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA	3319
TAGT	3323

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 851 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
Met Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln
 1             5             10             15

Lys Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp
      20             25             30

Leu Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu
      35             40             45

Ile Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser
      50             55             60

Thr Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val
      65             70             75             80

Leu Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr
      85             90             95

Leu Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu
      100            105            110

Val Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val
      115            120            125

Leu Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys
      130            135            140

Val Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu
      145            150            155            160

Pro Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro
      165            170            175

Arg Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn
      180            185            190

Asp Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile
      195            200            205

Asp Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn
      210            215            220
```

Ser Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu
 225 230 235 240
 Ser Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala
 245 250 255
 Arg Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp
 260 265 270
 Ser Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu
 275 280 285
 Val Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr
 290 295 300
 Ile Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser
 305 310 315 320
 Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu
 325 330 335
 Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys
 340 345 350
 Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg
 355 360 365
 Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu
 370 375 380
 Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu
 385 390 395 400
 Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val
 405 410 415
 Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly
 420 425 430
 Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala
 435 440 445
 Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn
 450 455 460
 Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile
 465 470 475 480
 Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile
 485 490 495
 Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala
 500 505 510

366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510

Asp	Ile	Glu	Gly	Ser	Asp	Glu	Ala	Asp	Gly	Ser	Lys	His	Leu	Pro	Gly	
				805					810					815		
Glu	Ser	Lys	Phe	Gln	Gln	Lys	Leu	Ala	Glu	Met	Thr	Ser	Thr	Arg	Thr	
			820					825					830			
Arg	Met	Gln	Lys	Gln	Lys	Met	Asn	Asp	Ser	Met	Asp	Thr	Ser	Asn	Lys	
		835					840					845				
Glu	Glu	Lys														
		850														

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3266 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 7..2502

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GCCATC ATG CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	48
Met Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
1 5 10	
GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	96
Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
15 20 25 30	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	144
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
35 40 45	
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG	192
Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu	
50 55 60	
TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG	240
Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu	
65 70 75	
ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT GCA TTG GTG	288
Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val	
80 85 90	
CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA	336
Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu	
95 100 105 110	

Biocompare.com

CAA	ATG	GAA	GAT	GAT	CTG	GTG	ATT	TCA	TTT	CAG	TTA	ATG	CTA	TGT	GTC	384
Gln	Met	Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	
					115				120					125		
CTT	GAC	TAT	TTT	ATT	AAA	CTC	TCA	CCT	CCC	ATG	TTG	CTC	AAA	GAA	CCA	432
Leu	Asp	Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	
			130					135					140			
TAT	AAA	ACA	GCT	GTT	ATA	CCC	ATT	AAT	GGT	TCA	CCT	CGA	ACA	CCC	AGG	480
Tyr	Lys	Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	
		145					150					155				
CGA	GGT	CAG	AAC	AGG	AGT	GCA	CGG	ATA	GCA	AAA	CAA	CTA	GAA	AAT	GAT	528
Arg	Gly	Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	
	160					165					170					
ACA	AGA	ATT	ATT	GAA	GTT	CTC	TGT	AAA	GAA	CAT	GAA	TGT	AAT	ATA	GAT	576
Thr	Arg	Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	
175					180				185					190		
GAG	GTG	AAA	AAT	GTT	TAT	TTC	AAA	AAT	TTT	ATA	CCT	TTT	ATG	AAT	TCT	624
Glu	Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	
			195					200					205			
CTT	GGA	CTT	GTA	ACA	TCT	AAT	GGA	CTT	CCA	GAG	GTT	GAA	AAT	CTT	TCT	672
Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	
			210					215				220				
AAA	CGA	TAC	GAA	GAA	ATT	TAT	CTT	AAA	AAT	AAA	GAT	CTA	GAT	GCA	AGA	720
Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	
		225					230					235				
TTA	TTT	TTG	GAT	CAT	GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	ATA	GAC	AGT	768
Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	
	240					245					250					
TTT	GAA	ACA	CAG	AGA	ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	GAA	GAG	GTG	816
Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	
255					260				265					270		
AAT	GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	AAC	ACT	ATC	864
Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	
			275					280				285				
CAA	CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	CCT	TCA	GAA	912
Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	
		290						295				300				
AAT	CTG	ATT	TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	AGT	960
Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	
	305					310						315				

ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT	1008
Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe	
320 325 330	
GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC	1056
Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr	
335 340 345 350	
AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA	1104
Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys	
355 360 365	
TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT	1152
Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn	
370 375 380	
GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA	1200
Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val	
385 390 395	
ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA	1248
Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr	
400 405 410	
GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT	1296
Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe	
415 420 425 430	
GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG	1344
Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu	
435 440 445	
ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG	1392
Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met	
450 455 460	
GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA	1440
Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys	
465 470 475	
CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT	1488
Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys	
480 485 490	
CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT	1536
Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr	
495 500 505 510	
CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA	1584
Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val	
515 520 525	

AAT Asn	TCT Ser	ACT Thr	GCA Ala	AAT Asn	GCA Ala	GAG Glu	ACA Thr	CAA Gln	GCA Ala	ACC Thr	TCA Ser	GCC Ala	TTC Phe	CAG Gln	ACC Thr	1632	
			530				535				540						
CAG Gln	AAG Lys	CCA Pro	TTG Leu	AAA Lys	TCT Ser	ACC Thr	TCT Ser	CTT Leu	TCA Ser	CTG Leu	TTT Phe	TAT Tyr	AAA Lys	AAA Lys	GTG Val	1680	
			545				550				555						
TAT Tyr	CGG Arg	CTA Leu	GCC Ala	TAT Tyr	CTC Leu	CGG Arg	CTA Leu	AAT Asn	ACA Thr	CTT Leu	TGT Cys	GAA Glu	CGC Arg	CTT Leu	CTG Leu	1728	
			560				565				570						
TCT Ser	GAG Glu	CAC His	CCA Pro	GAA Glu	TTA Leu	GAA Glu	CAT His	ATC Ile	ATC Ile	TGG Trp	ACC Thr	CTT Leu	TTC Phe	CAG Gln	CAC His	1776	
			575				580				585					590	
ACC Thr	CTG Leu	CAG Gln	AAT Asn	GAG Glu	TAT Tyr	GAA Glu	CTC Leu	ATG Met	AGA Arg	GAC Asp	AGG Arg	CAT His	TTG Leu	GAC Asp	CAA Gln	1824	
			595				600				605						
ATT Ile	ATG Met	ATG Met	TGT Cys	TCC Ser	ATG Met	TAT Tyr	GGC Gly	ATA Ile	TGC Cys	AAA Lys	GTG Val	AAG Lys	AAT Asn	ATA Ile	GAC Asp	1872	
			610				615				620						
CTT Leu	AAA Lys	TTC Phe	AAA Lys	ATC Ile	ATT Ile	GTA Val	ACA Thr	GCA Ala	TAC Tyr	AAG Lys	GAT Asp	CTT Leu	CCT Pro	CAT His	GCT Ala	1920	
			625				630				635						
GTT Val	CAG Gln	GAG Glu	ACA Thr	TTC Phe	AAA Lys	CGT Arg	GTT Val	TTG Leu	ATC Ile	AAA Lys	GAA Glu	GAG Glu	GAG Glu	TAT Tyr	GAT Asp	1968	
			640				645				650						
TCT Ser	ATT Ile	ATA Ile	GTA Val	TTC Phe	TAT Tyr	AAC Asn	TCG Ser	GTC Val	TTC Phe	ATG Met	CAG Gln	AGA Arg	CTG Leu	AAA Lys	ACA Thr	2016	
			655				660				665					670	
AAT Asn	ATT Ile	TTG Leu	CAG Gln	TAT Tyr	GCT Ala	TCC Ser	ACC Thr	AGG Arg	CCC Pro	CCT Pro	ACC Thr	TTG Leu	TCA Ser	CCA Pro	ATA Ile	2064	
			675				680				685						
CCT Pro	CAC His	ATT Ile	CCT Pro	CGA Arg	AGC Ser	CCT Pro	TAC Tyr	AAG Lys	TTT Phe	CCT Pro	AGT Ser	TCA Ser	CCC Pro	TTA Leu	CGG Arg	2112	
			690				695				700						
ATT Ile	CCT Pro	GGA Gly	GGG Gly	AAC Asn	ATC Ile	TAT Tyr	ATT Ile	TCA Ser	CCC Pro	CTG Leu	AAG Lys	AGT Ser	CCA Pro	TAT Tyr	AAA Lys	2160	
			705				710				715						
ATT Ile	TCA Ser	GAA Glu	GGT Gly	CTG Leu	CCA Pro	ACA Thr	CCA Pro	ACA Thr	AAA Lys	ATG Met	ACT Thr	CCA Pro	AGA Arg	TCA Ser	AGA Arg	2208	
			720				725				730						

ATC Ile	TTA Leu	GTA Val	TCA Ser	ATT Ile	GGT Gly	GAA Glu	TCA Ser	TTC Phe	GGG Gly	ACT Thr	TCT Ser	GAG Glu	AAG Lys	TTC Phe	CAG Gln	2256	
					740					745					750		
AAA Lys	ATA Ile	AAT Asn	CAG Gln	ATG Met	GTA Val	TGT Cys	AAC Asn	AGC Ser	GAC Asp	CGT Arg	GTG Val	CTC Leu	AAA Lys	AGA Arg	AGT Ser	2304	
				755					760					765			
GCT Ala	GAA Glu	GGA Gly	AGC Ser	AAC Asn	CCT Pro	CCT Pro	AAA Lys	CCA Pro	CTG Leu	AAA Lys	AAA Lys	CTA Leu	CGC Arg	TTT Phe	GAT Asp	2352	
				770					775					780			
ATT Ile	GAA Glu	GGA Gly	TCA Ser	GAT Asp	GAA Glu	GCA Ala	GAT Asp	GGA Gly	AGT Ser	AAA Lys	CAT His	CTC Leu	CCA Pro	GGA Gly	GAG Glu	2400	
				785					790					795			
TCC Ser	AAA Lys	TTT Phe	CAG Gln	CAG Gln	AAA Lys	CTG Leu	GCA Ala	GAA Glu	ATG Met	ACT Thr	TCT Ser	ACT Thr	CGA Arg	ACA Thr	CGA Arg	2448	
				800					805					810			
ATG Met	CAA Gln	AAG Lys	CAG Gln	AAA Lys	ATG Met	AAT Asn	GAT Asp	AGC Ser	ATG Met	GAT Asp	ACC Thr	TCA Ser	AAC Asn	AAG Lys	GAA Glu	2496	
				815					820					825			830
GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG Glu Lys																2552	
TCTCTCACAG ATGTGACTGT ATAACCTTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC																2612	
TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC																2672	
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTG TGTA AATCCTGCCA																2732	
TTTAAAAAGT TGTAGCAGAT TGTTTTCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT																2792	
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT																2852	
TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT																2912	
TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCCTATCT ATCTTCCAAA TGCAATTTGA																2972	
TTGACTGCCC ATTCAACCAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA																3032	
ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT																3092	
ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT																3152	
AACCATATGA TACTATCATA CTACTGAAAC AGATTTTATA CCTCAGAATG TAAAAGAACT																3212	
TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT																3266	

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 832 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu
1				5					10					15	
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val
			20					25					30		
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val
		35					40					45			
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala
	50					55					60				
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln
65					70					75					80
Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys
				85					90					95	
Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	Met
			100					105					110		
Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp
	115						120					125			
Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys
	130					135					140				
Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly
145					150					155					160
Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	Arg
			165						170					175	
Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu	Val
		180						185					190		
Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu	Gly
	195						200					205			
Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	Ser	Lys	Arg
	210					215					220				

Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	Arg	Leu	Phe	225	230	235	240
Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	Ser	Phe	Glu	245	250	255	
Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	Val	Asn	Val	260	265	270	
Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	Ile	Gln	Gln	275	280	285	
Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	Glu	Asn	Leu	290	295	300	
Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu	305	310	315	320
Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys	325	330	335	
Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	340	345	350	
Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu	355	360	365	
Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	Asn	370	375	380	
Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	385	390	395	400
Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	405	410	415	
Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	420	425	430	
Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	435	440	445	
Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	450	455	460	
Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	465	470	475	480
Lys	Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	485	490	495	
Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	500	505	510	

Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	515	520	525
Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	530	535	540
Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	545	550	555
Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	565	570	575
His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	580	585	590
Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	595	600	605
Met	Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	610	615	620
Phe	Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	625	630	635
Glu	Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	645	650	655
Ile	Val	Phe	Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	660	665	670
Leu	Gln	Tyr	Ala	Ser	Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	675	680	685
Ile	Pro	Arg	Ser	Pro	Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	690	695	700
Gly	Gly	Asn	Ile	Tyr	Ile	Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	705	710	715
Glu	Gly	Leu	Pro	Thr	Pro	Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu	725	730	735
Val	Ser	Ile	Gly	Glu	Ser	Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile	740	745	750
Asn	Gln	Met	Val	Cys	Asn	Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu	755	760	765
Gly	Ser	Asn	Pro	Pro	Lys	Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu	770	775	780
Gly	Ser	Asp	Glu	Ala	Asp	Gly	Ser	Lys	His	Leu	Pro	Gly	Glu	Ser	Lys	785	790	795

Phe	Gln	Gln	Lys	Leu	Ala	Glu	Met	Thr	Ser	Thr	Arg	Thr	Arg	Met	Gln
				805					810					815	
Lys	Gln	Lys	Met	Asn	Asp	Ser	Met	Asp	Thr	Ser	Asn	Lys	Glu	Glu	Lys
			820					825					830		

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3113 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2349

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GCCGTC	ATG	TCA	AGA	CTG	TTG	AAG	AAG	TAT	GAT	GTA	TTG	TTT	GCA	CTC	48
Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	Leu		
1				5					10						
TTC	AGC	AAA	TTG	GAA	AGG	ACA	TGT	GAA	CTT	ATA	TAT	TTG	ACA	CAA	CCC
Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	Pro
15				20				25						30	
AGC	AGT	TCG	ATA	TCT	ACT	GAA	ATA	AAT	TCT	GCA	TTG	GTG	CTA	AAA	GTT
Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys	Val
			35					40						45	
TCT	TGG	ATC	ACA	TTT	TTA	TTA	GCT	AAA	GGG	GAA	GTA	TTA	CAA	ATG	GAA
Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	Met	Glu
			50					55					60		
GAT	GAT	CTG	GTG	ATT	TCA	TTT	CAG	TTA	ATG	CTA	TGT	GTC	CTT	GAC	TAT
Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp	Tyr
		65					70					75			
TTT	ATT	AAA	CTC	TCA	CCT	CCC	ATG	TTG	CTC	AAA	GAA	CCA	TAT	AAA	ACA
Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys	Thr
		80					85				90				
GCT	GTT	ATA	CCC	ATT	AAT	GGT	TCA	CCT	CGA	ACA	CCC	AGG	CGA	GGT	CAG
Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly	Gln
		95				100				105				110	
AAC	AGG	AGT	GCA	CGG	ATA	GCA	AAA	CAA	CTA	GAA	AAT	GAT	ACA	AGA	ATT
Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	Arg	Ile
			115						120					125	

ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT GAG GTG AAA	432
Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys	
130 135 140	
AAT GTT TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT CTT GGA CTT	480
Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu	
145 150 155	
GTA ACA TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT AAA CGA TAC	528
Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr	
160 165 170	
GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA TTA TTT TTG	576
Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu	
175 180 185 190	
GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT TTT GAA ACA	624
Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr	
195 200 205	
CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG AAT GTA ATT	672
Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile	
210 215 220	
CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC CAA CAA TTA	720
Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu	
225 230 235	
ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA AAT CTG ATT	768
Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile	
240 245 250	
TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT ATA CTG AAA	816
Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys	
255 260 265 270	
AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT GCT AAA GCT	864
Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala	
275 280 285	
GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC AAA CTT GGA	912
Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly	
290 295 300	
GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA TCA GAA GAA	960
Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu	
305 310 315	
GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT GAC AAC ATT	1008
Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile	
320 325 330	

TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA ATG GCC ACA	1056
Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr	
335 340 345 350	
TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA GAT TTG TCT	1104
Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser	
355 360 365	
TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT GAT TTT TAC	1152
Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr	
370 375 380	
AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG ACA AGA GAA	1200
Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu	
385 390 395	
ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG GAA TCC CTT	1248
Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu	
400 405 410	
GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA CAA TCA AAG	1296
Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys	
415 420 425 430	
GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT CCT CTT AAT	1344
Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn	
435 440 445	
CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT CTT TCT CCT	1392
Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro	
450 455 460	
GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT	1440
Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr	
465 470 475	
GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA	1488
Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro	
480 485 490	
TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA	1536
Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu	
495 500 505 510	
GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG TCT GAG CAC	1584
Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His	
515 520 525	
CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC ACC CTG CAG	1632
Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln	
530 535 540	

AAT Asn	GAG Glu	TAT Tyr	GAA Glu	CTC Leu	ATG Met	AGA Arg	GAC Asp	AGG Arg	CAT His	TTG Leu	GAC Asp	CAA Gln	ATT Ile	ATG Met	ATG Met	1680				
545																550	555			
TGT Cys	TCC Ser	ATG Met	TAT Tyr	GGC Gly	ATA Ile	TGC Cys	AAA Lys	GTG Val	AAG Lys	AAT Asn	ATA Ile	GAC Asp	CTT Leu	AAA Lys	TTC Phe	1728				
560																565	570			
AAA Lys	ATC Ile	ATT Ile	GTA Val	ACA Thr	GCA Ala	TAC Tyr	AAG Lys	GAT Asp	CTT Leu	CCT Pro	CAT His	GCT Ala	GTT Val	CAG Gln	GAG Glu	1776				
575																580	585			590
ACA Thr	TTC Phe	AAA Lys	CGT Arg	GTT Val	TTG Leu	ATC Ile	AAA Lys	GAA Glu	GAG Glu	GAG Glu	TAT Tyr	GAT Asp	TCT Ser	ATT Ile	ATA Ile	1824				
595																600	605			
GTA Val	TTC Phe	TAT Tyr	AAC Asn	TCG Ser	GTC Val	TTC Phe	ATG Met	CAG Gln	AGA Arg	CTG Leu	AAA Lys	ACA Thr	AAT Asn	ATT Ile	TTG Leu	1872				
610																615	620			
CAG Gln	TAT Tyr	GCT Ala	TCC Ser	ACC Thr	AGG Arg	CCC Pro	CCT Pro	ACC Thr	TTG Leu	TCA Ser	CCA Pro	ATA Ile	CCT Pro	CAC His	ATT Ile	1920				
625																630	635			
CCT Pro	CGA Arg	AGC Ser	CCT Pro	TAC Tyr	AAG Lys	TTT Phe	CCT Pro	AGT Ser	TCA Ser	CCC Pro	TTA Leu	CGG Arg	ATT Ile	CCT Pro	GGA Gly	1968				
640																645	650			
GGG Gly	AAC Asn	ATC Ile	TAT Tyr	ATT Ile	TCA Ser	CCC Pro	CTG Leu	AAG Lys	AGT Ser	CCA Pro	TAT Tyr	AAA Lys	ATT Ile	TCA Ser	GAA Glu	2016				
655																660	665			670
GGT Gly	CTG Leu	CCA Pro	ACA Thr	CCA Pro	ACA Thr	AAA Lys	ATG Met	ACT Thr	CCA Pro	AGA Arg	TCA Ser	AGA Arg	ATC Ile	TTA Leu	GTA Val	2064				
675																680	685			
TCA Ser	ATT Ile	GGT Gly	GAA Glu	TCA Ser	TTC Phe	GGG Gly	ACT Thr	TCT Ser	GAG Glu	AAG Lys	TTC Phe	CAG Gln	AAA Lys	ATA Ile	AAT Asn	2112				
690																695	700			
CAG Gln	ATG Met	GTA Val	TGT Cys	AAC Asn	AGC Ser	GAC Asp	CGT Arg	GTG Val	CTC Leu	AAA Lys	AGA Arg	AGT Ser	GCT Ala	GAA Glu	GGA Gly	2160				
705																710	715			
AGC Ser	AAC Asn	CCT Pro	CCT Pro	AAA Lys	CCA Pro	CTG Leu	AAA Lys	AAA Lys	CTA Leu	CGC Arg	TTT Phe	GAT Asp	ATT Ile	GAA Glu	GGA Gly	2208				
720																725	730			
TCA Ser	GAT Asp	GAA Glu	GCA Ala	GAT Asp	GGA Gly	AGT Ser	AAA Lys	CAT His	CTC Leu	CCA Pro	GGA Gly	GAG Glu	TCC Ser	AAA Lys	TTT Phe	2256				
735																740	745			750

CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA ATG CAA AAG	2304
Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys	
755 760 765	
CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG AAA	2349
Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys	
770 775 780	
TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG TCTCTCACAG	2409
ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT	2469
TTTGTGGATA TAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG	2529
TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA TTTAAAAAGT	2589
TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG	2649
GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG	2709
CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC	2769
ATGAACACCC TTAGAAAATG TGTCCCTATCT ATCTTCCAAA TGCAATTGA TTGACTGCCC	2829
ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA	2889
AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT ACTGTGTGCT	2949
TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA	3009
TACTATCATA CTACTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT	3069
TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT	3113

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 781 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala Leu Phe Ser	
1 5 10 15	
Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln Pro Ser Ser	
20 25 30	
Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val Ser Trp	
35 40 45	

Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	Thr	Tyr	Ser	340	345	350	
Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	Ser	Phe	Pro	355	360	365	
Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	Tyr	Lys	Val	370	375	380	
Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	Glu	Met	Ile	385	390	395	400
Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	Leu	Ala	Trp	405	410	415	
Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	Lys	Asp	Arg	420	425	430	
Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	Asn	Leu	Pro	435	440	445	
Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	Pro	Val	Arg	450	455	460	
Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	Thr	Ala	Asn	465	470	475	480
Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	Pro	Leu	Lys	485	490	495	
Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	Leu	Ala	Tyr	500	505	510	
Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	His	Pro	Glu	515	520	525	
Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	Gln	Asn	Glu	530	535	540	
Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	Met	Cys	Ser	545	550	555	560
Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	Phe	Lys	Ile	565	570	575	
Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	Glu	Thr	Phe	580	585	590	
Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	Ile	Val	Phe	595	600	605	
Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	Leu	Gln	Tyr	610	615	620	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Ala	Ser	Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	Ile	Pro	Arg	
625					630					635					640	
Ser	Pro	Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	Gly	Gly	Asn	
				645					650					655		
Ile	Tyr	Ile	Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	Glu	Gly	Leu	
			660					665					670			
Pro	Thr	Pro	Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu	Val	Ser	Ile	
			675				680					685				
Gly	Glu	Ser	Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile	Asn	Gln	Met	
	690					695				700						
Val	Cys	Asn	Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu	Gly	Ser	Asn	
705					710					715					720	
Pro	Pro	Lys	Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu	Gly	Ser	Asp	
				725					730					735		
Glu	Ala	Asp	Gly	Ser	Lys	His	Leu	Pro	Gly	Glu	Ser	Lys	Phe	Gln	Gln	
			740					745					750			
Lys	Leu	Ala	Glu	Met	Thr	Ser	Thr	Arg	Thr	Arg	Met	Gln	Lys	Gln	Lys	
			755				760					765				
Met	Asn	Asp	Ser	Met	Asp	Thr	Ser	Asn	Lys	Glu	Glu	Lys				
	770					775					780					

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3323 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2559

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CGCGTC	ATG	CCG	CCC	AAA	ACC	CCC	CGA	AAA	ACG	GCC	GCC	ACC	GCC	GCC	48
Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala		
1					5					10					
GCT	GCC	GCC	GCG	GAA	CCC	CCG	GCA	CCG	CCG	CCG	CCC	CCT	CCT	GAG	96
Ala	Ala	Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Glu	
15					20					25				30	

GTC Val	GAC Asp	CTA Leu	GAT Asp	GAG Glu	ATG Met	TCG Ser	TTC Phe	ACT Thr	TTT Phe	ACT Thr	GAG Glu	CTA Leu	CAG Gln	AAA Lys	AAC Asn	144
				35					40					45		
ATA Ile	GAA Glu	ATC Ile	AGT Ser	GTC Val	CAT His	AAA Lys	TTC Phe	TTT Phe	AAC Asn	TTA Leu	CTA Leu	AAA Lys	GAA Glu	ATT Ile	GAT Asp	192
				50					55					60		
ACC Thr	AGT Ser	ACC Thr	AAA Lys	GTT Val	GAT Asp	AAT Asn	GCT Ala	ATG Met	TCA Ser	AGA Arg	CTG Leu	TTG Leu	AAG Lys	AAG Lys	TAT Tyr	240
				65					70					75		
GAT Asp	GTA Val	TTG Leu	TTT Phe	GCA Ala	CTC Leu	TTC Phe	AGC Ser	AAA Lys	TTG Leu	GAA Glu	AGG Arg	ACA Thr	TGT Cys	GAA Glu	CTT Leu	288
				80					85					90		
ATA Ile	TAT Tyr	TTG Leu	ACA Thr	CAA Gln	CCC Pro	AGC Ser	AGT Ser	TCG Ser	ATA Ile	TCT Ser	ACT Thr	GAA Glu	ATA Ile	AAT Asn	TCT Ser	336
				95					100					105		
GCA Ala	TTG Leu	GTG Val	CTA Leu	AAA Lys	GTT Val	TCT Ser	TGG Trp	ATC Ile	ACA Thr	TTT Phe	TTA Leu	TTA Leu	GCT Ala	AAA Lys	GGG Gly	384
				115					120					125		
GAA Glu	GTA Val	TTA Leu	CAA Gln	ATG Met	GAA Glu	GAT Asp	GAT Asp	CTG Leu	GTG Val	ATT Ile	TCA Ser	TTT Phe	CAG Gln	TTA Leu	ATG Met	432
				130					135					140		
CTA Leu	TGT Cys	GTC Val	CTT Leu	GAC Asp	TAT Tyr	TTT Phe	ATT Ile	AAA Lys	CTC Leu	TCA Ser	CCT Pro	CCC Pro	ATG Met	TTG Leu	CTC Leu	480
				145					150					155		
AAA Lys	GAA Glu	CCA Pro	TAT Tyr	AAA Lys	ACA Thr	GCT Ala	GTT Val	ATA Ile	CCC Pro	ATT Ile	AAT Asn	GGT Gly	TCA Ser	CCT Pro	CGA Arg	528
				160					165					170		
ACA Thr	CCC Pro	AGG Arg	CGA Arg	GGT Gly	CAG Gln	AAC Asn	AGG Arg	AGT Ser	GCA Ala	CGG Arg	ATA Ile	GCA Ala	AAA Lys	CAA Gln	CTA Leu	576
				175					180					185		
GAA Glu	AAT Asn	GAT Asp	ACA Thr	AGA Arg	ATT Ile	ATT Ile	GAA Glu	GTT Val	CTC Leu	TGT Cys	AAA Lys	GAA Glu	CAT His	GAA Glu	TGT Cys	624
				195					200					205		
AAT Asn	ATA Ile	GAT Asp	GAG Glu	GTG Val	AAA Lys	AAT Asn	GTT Val	TAT Tyr	TTC Phe	AAA Lys	AAT Asn	TTT Phe	ATA Ile	CCT Pro	TTT Phe	672
				210					215					220		
ATG Met	AAT Asn	TCT Ser	CTT Leu	GGA Gly	CTT Leu	GTA Val	ACA Thr	TCT Ser	AAT Asn	GGA Gly	CTT Leu	CCA Pro	GAG Glu	GTT Val	GAA Glu	720
				225					230					235		

AAT CTT TCT AAA CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA	768
Asn Leu Ser Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu	
240 245 250	
GAT GCA AGA TTA TTT TTG GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT	816
Asp Ala Arg Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser	
255 260 265 270	
ATA GAC AGT TTT GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT	864
Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp	
275 280 285	
GAA GAG GTG AAT GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG	912
Glu Glu Val Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met	
290 295 300	
AAC ACT ATC CAA CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA	960
Asn Thr Ile Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln	
305 310 315	
CCT TCA GAA AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA	1008
Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro	
320 325 330	
AAA GAA AGT ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA	1056
Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys	
335 340 345 350	
GAG AAA TTT GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA	1104
Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser	
355 360 365	
CAG CGA TAC AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC	1152
Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser	
370 375 380	
ATG CTT AAA TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA	1200
Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys	
385 390 395	
CTT CTG AAT GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT	1248
Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu	
400 405 410	
GAG GTT GTA ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT	1296
Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp	
415 420 425 430	
TCT GGA ACA GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA	1344
Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu	
435 440 445	

AAA	GCC	TTT	GAT	TTT	TAC	AAA	GTG	ATC	GAA	AGT	TTT	ATC	AAA	GCA	GAA	1392
Lys	Ala	Phe	Asp	Phe	Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	
			450						455					460		
GGC	AAC	TTG	ACA	AGA	GAA	ATG	ATA	AAA	CAT	TTA	GAA	CGA	TGT	GAA	CAT	1440
Gly	Asn	Leu	Thr	Arg	Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	
			465					470					475			
CGA	ATC	ATG	GAA	TCC	CTT	GCA	TGG	CTC	TCA	GAT	TCA	CCT	TTA	TTT	GAT	1488
Arg	Ile	Met	Glu	Ser	Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	
			480				485				490					
CTT	ATT	AAA	CAA	TCA	AAG	GAC	CGA	GAA	GGA	CCA	ACT	GAT	CAC	CTT	GAA	1536
Leu	Ile	Lys	Gln	Ser	Lys	Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	
						500				505					510	
TCT	GCT	TGT	CCT	CTT	AAT	CTT	CCT	CTC	CAG	AAT	AAT	CAC	ACT	GCA	GCA	1584
Ser	Ala	Cys	Pro	Leu	Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	
					515				520					525		
GAT	ATG	TAT	CTT	TCT	CCT	GTA	AGA	TCT	CCA	AAG	AAA	AAA	GGT	TCA	ACT	1632
Asp	Met	Tyr	Leu	Ser	Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	
			530					535					540			
ACG	CGT	GTA	AAT	TCT	ACT	GCA	AAT	GCA	GAG	ACA	CAA	GCA	ACC	TCA	GCC	1680
Thr	Arg	Val	Asn	Ser	Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	
			545					550				555				
TTC	CAG	ACC	CAG	AAG	CCA	TTG	AAA	TCT	ACC	TCT	CTT	TCA	CTG	TTT	TAT	1728
Phe	Gln	Thr	Gln	Lys	Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	
			560				565				570					
AAA	AAA	GTG	TAT	CGG	CTA	GCC	TAT	CTC	CGG	CTA	AAT	ACA	CTT	TGT	GAA	1776
Lys	Lys	Val	Tyr	Arg	Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	
						580				585					590	
CGC	CTT	CTG	TCT	GAG	CAC	CCA	GAA	TTA	GAA	CAT	ATC	ATC	TGG	ACC	CTT	1824
Arg	Leu	Leu	Ser	Glu	His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	
					595				600					605		
TTC	CAG	CAC	ACC	CTG	CAG	AAT	GAG	TAT	GAA	CTC	ATG	AGA	GAC	AGG	CAT	1872
Phe	Gln	His	Thr	Leu	Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	
			610					615					620			
TTG	GAC	CAA	ATT	ATG	ATG	TGT	TCC	ATG	TAT	GGC	ATA	TGC	AAA	GTG	AAG	1920
Leu	Asp	Gln	Ile	Met	Met	Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	
			625					630				635				
AAT	ATA	GAC	CTT	AAA	TTC	AAA	ATC	ATT	GTA	ACA	GCA	TAC	AAG	GAT	CTT	1968
Asn	Ile	Asp	Leu	Lys	Phe	Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	
			640				645					650				

CCT CAT GCT GTT CAG GAG ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG	2016
Pro His Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu	
655 660 665 670	
GAG TAT GAT TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA	2064
Glu Tyr Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg	
675 680 685	
CTG AAA ACA AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG	2112
Leu Lys Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu	
690 695 700	
TCA CCA ATA CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA	2160
Ser Pro Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser	
705 710 715	
CCC TTA CGG ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT	2208
Pro Leu Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser	
720 725 730	
CCA TAT AAA ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA	2256
Pro Tyr Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro	
735 740 745 750	
AGA TCA AGA ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG	2304
Arg Ser Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu	
755 760 765	
AAG TTC CAG AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC	2352
Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu	
770 775 780	
AAA AGA AGT GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA	2400
Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu	
785 790 795	
CGC TTT GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC	2448
Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu	
800 805 810	
CCA GGA GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT	2496
Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr	
815 820 825 830	
CGA ACA CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA	2544
Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser	
835 840 845	
AAC AAG GAA GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT	2599
Asn Lys Glu Glu Lys	
850	
GGATTCATTG TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC	2659

ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA 2719
 TTCCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA 2779
 AATCCTGCCA TTTAAAAAGT TGTAGCAGAT TGTTCCTCT TCCAAAGTAA AATTGCTGTG 2839
 CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC 2899
 TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT 2959
 GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA 3019
 TGCAATTTGA TTGACTGCCC ATTCACCAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT 3079
 ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG 3139
 AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA 3199
 AGCAAAGTAT AACCATATGA TACTATCATA CTACTGAAAC AGATTTTCATA CCTCAGAAATG 3259
 TAAAAGAAGT TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA 3319
 TAGT 3323

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 851 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	1	5	10	15
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Val	Asp	20	25	30	
Leu	Asp	Glu	Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	35	40	45	
Ile	Ser	Val	His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	50	55	60	
Thr	Lys	Val	Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	65	70	75	80
Leu	Phe	Ala	Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	85	90	95	

Leu Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu
 100 105 110
 Val Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val
 115 120 125
 Leu Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys
 130 135 140
 Val Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu
 145 150 155 160
 Pro Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro
 165 170 175
 Arg Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn
 180 185 190
 Asp Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile
 195 200 205
 Asp Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn
 210 215 220
 Ser Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu
 225 230 235 240
 Ser Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala
 245 250 255
 Arg Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp
 260 265 270
 Ser Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu
 275 280 285
 Val Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr
 290 295 300
 Ile Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser
 305 310 315 320
 Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu
 325 330 335
 Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys
 340 345 350
 Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg
 355 360 365
 Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu
 370 375 380

Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys
 675 680 685
 Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro
 690 695 700
 Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu
 705 710 715 720
 Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr
 725 730 735
 Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser
 740 745 750
 Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe
 755 760 765
 Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg
 770 775 780
 Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe
 785 790 795 800
 Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly
 805 810 815
 Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr
 820 825 830
 Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys
 835 840 845
 Glu Glu Lys
 850

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3461 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2697

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC
 Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala
 1 5 10

GCT	GCC	GCG	GAA	CCC	CCG	GCA	CCG	CCG	CCG	CCG	CCC	CCT	CCT	GAG		96
Ala	Ala	Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Glu		
15				20					25					30		
GAG	GAC	CCA	GAG	CAG	GAC	AGC	GGC	CCG	GAG	GAC	CTG	CCT	CTC	GTC	AGG	144
Glu	Asp	Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	
			35					40						45		
CTT	GAG	TTT	GAA	GAA	ACA	GAA	GAA	CCT	GAT	TTT	ACT	GCA	TTA	TGT	CAG	192
Leu	Glu	Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	
			50					55					60			
AAA	TTA	AAG	ATA	CCA	GAT	CAT	GTC	AGA	GAG	AGA	GCT	TGG	TTG	GTC	GAC	240
Lys	Leu	Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Val	Asp	
		65					70					75				
CTA	GAT	GAG	ATG	TCG	TTC	ACT	TTT	ACT	GAG	CTA	CAG	AAA	AAC	ATA	GAA	288
Leu	Asp	Glu	Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	
	80					85					90					
ATC	AGT	GTC	CAT	AAA	TTC	TTT	AAC	TTA	CTA	AAA	GAA	ATT	GAT	ACC	AGT	336
Ile	Ser	Val	His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	
	95			100					105						110	
ACC	AAA	GTT	GAT	AAT	GCT	ATG	TCA	AGA	CTG	TTG	AAG	AAG	TAT	GAT	GTA	384
Thr	Lys	Val	Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	
			115					120					125			
TTG	TTT	GCA	CTC	TTC	AGC	AAA	TTG	GAA	AGG	ACA	TGT	GAA	CTT	ATA	TAT	432
Leu	Phe	Ala	Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	
		130						135					140			
TTG	ACA	CAA	CCC	AGC	AGT	TCG	ATA	TCT	ACT	GAA	ATA	AAT	TCT	GCA	TTG	480
Leu	Thr	Gln	Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	
		145					150					155				
GTG	CTA	AAA	GTT	TCT	TGG	ATC	ACA	TTT	TTA	TTA	GCT	AAA	GGG	GAA	GTA	528
Val	Leu	Lys	Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	
	160					165					170					
TTA	CAA	ATG	GAA	GAT	GAT	CTG	GTG	ATT	TCA	TTT	CAG	TTA	ATG	CTA	TGT	576
Leu	Gln	Met	Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	
175				180					185						190	
GTC	CTT	GAC	TAT	TTT	ATT	AAA	CTC	TCA	CCT	CCC	ATG	TTG	CTC	AAA	GAA	624
Val	Leu	Asp	Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	
			195					200						205		
CCA	TAT	AAA	ACA	GCT	GTT	ATA	CCC	ATT	AAT	GGT	TCA	CCT	CGA	ACA	CCC	672
Pro	Tyr	Lys	Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	
		210					215						220			

AGG	CGA	GGT	CAG	AAC	AGG	AGT	GCA	CGG	ATA	GCA	AAA	CAA	CTA	GAA	AAT	720
Arg	Arg	Gly	Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	
		225						230				235				
GAT	ACA	AGA	ATT	ATT	GAA	GTT	CTC	TGT	AAA	GAA	CAT	GAA	TGT	AAT	ATA	768
Asp	Thr	Arg	Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	
		240						245				250				
GAT	GAG	GTG	AAA	AAT	GTT	TAT	TTC	AAA	AAT	TTT	ATA	CCT	TTT	ATG	AAT	816
Asp	Glu	Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	
		255						260				265			270	
TCT	CTT	GGA	CTT	GTA	ACA	TCT	AAT	GGA	CTT	CCA	GAG	GTT	GAA	AAT	CTT	864
Ser	Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	Asn	Leu	
		275						280				285				
TCT	AAA	CGA	TAC	GAA	GAA	ATT	TAT	CTT	AAA	AAT	AAA	GAT	CTA	GAT	GCA	912
Ser	Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	Asp	Ala	
		290						295				300				
AGA	TTA	TTT	TTG	GAT	CAT	GAT	AAA	ACT	CTT	CAG	ACT	GAT	TCT	ATA	GAC	960
Arg	Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	Ile	Asp	
		305						310				315				
AGT	TTT	GAA	ACA	CAG	AGA	ACA	CCA	CGA	AAA	AGT	AAC	CTT	GAT	GAA	GAG	1008
Ser	Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	Glu	Glu	
		320						325				330				
GTG	AAT	GTA	ATT	CCT	CCA	CAC	ACT	CCA	GTT	AGG	ACT	GTT	ATG	AAC	ACT	1056
Val	Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	Asn	Thr	
		335						340				345			350	
ATC	CAA	CAA	TTA	ATG	ATG	ATT	TTA	AAT	TCA	GCA	AGT	GAT	CAA	CCT	TCA	1104
Ile	Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	Pro	Ser	
		355						360				365				
GAA	AAT	CTG	ATT	TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	1152
Glu	Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	
		370						375				380				
AGT	ATA	CTG	AAA	AGA	GTG	AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	GAG	AAA	1200
Ser	Ile	Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	
		385						390				395				
TTT	GCT	AAA	GCT	GTG	GGA	CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	CAG	CGA	1248
Phe	Ala	Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	
		400						405				410				
TAC	AAA	CTT	GGA	GTT	CGC	TTG	TAT	TAC	CGA	GTA	ATG	GAA	TCC	ATG	CTT	1296
Tyr	Lys	Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	
		415						420				425			430	

AAA TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG	1344
Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu	
435 440 445	
AAT GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT	1392
Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val	
450 455 460	
GTA ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA	1440
Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly	
465 470 475	
ACA GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC	1488
Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala	
480 485 490	
TTT GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC	1536
Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn	
495 500 505 510	
TTG ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC	1584
Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile	
515 520 525	
ATG GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT	1632
Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile	
530 535 540	
AAA CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT	1680
Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala	
545 550 555	
TGT CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG	1728
Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met	
560 565 570	
TAT CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT	1776
Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg	
575 580 585 590	
GTA AAT TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG	1824
Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln	
595 600 605	
ACC CAG AAG CCA TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA	1872
Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys	
610 615 620	
GTG TAT CGG CTA GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT	1920
Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu	
625 630 635	

CTG TCT GAG CAC CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG	1968
Leu Ser Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln	
640 645 650	
CAC ACC CTG CAG AAT GAG TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC	2016
His Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp	
655 660 665 670	
CAA ATT ATG ATG TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA	2064
Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile	
675 680 685	
GAC CTT AAA TTC AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT	2112
Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His	
690 695 700	
GCT GTT CAG GAG ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT	2160
Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr	
705 710 715	
GAT TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA	2208
Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys	
720 725 730	
ACA AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA	2256
Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro	
735 740 745 750	
ATA CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA	2304
Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu	
755 760 765	
CGG ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT	2352
Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr	
770 775 780	
AAA ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA	2400
Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser	
785 790 795	
AGA ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC	2448
Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe	
800 805 810	
CAG AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA	2496
Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg	
815 820 825 830	
AGT GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT	2544
Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe	
835 840 845	

GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA	2592
Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly	
850 855 860	
GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA	2640
Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr	
865 870 875	
CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG	2688
Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys	
880 885 890	
GAA GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT	2737
Glu Glu Lys	
895	
GGATTTCATTG TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC	2797
ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAATGTGC AGATGCAATT GTTTGGGTGA	2857
TTCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA	2917
AATCCTGCCA TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG	2977
CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC	3037
TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT	3097
GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA	3157
TGCAATTTGA TTGACTGCCC ATTCACCAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT	3217
ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG	3277
AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA	3337
AGCAAAGTAT AACCATATGA TACTATCATA CTACTGAAAC AGATTTTATA CCTCAGAATG	3397
TAAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA	3457
TAGT	3461

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 897 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	
1				5					10					15		
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp	
			20					25					30			
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	
			35				40					45				
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	
	50					55					60					
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Val	Asp	Leu	Asp	
65					70				75					80		
Glu	Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	
				85					90					95		
Val	His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	
			100					105					110			
Val	Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	
		115					120					125				
Ala	Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	
	130					135					140					
Gln	Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	
145					150					155					160	
Lys	Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	
				165					170					175		
Met	Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	
			180					185					190			
Asp	Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	
		195					200					205				
Lys	Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	
	210					215					220					
Gly	Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	
225					230					235				240		
Arg	Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu	
				245				250						255		
Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu	
			260					265					270			

AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GTC GAA TCT ACT GAA ATA AAT TCT GCA TTG GTG CTA AAA GTT TCT TGG	384
Val Glu Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys Val Ser Trp	
115 120 125	
ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA CAA ATG GAA GAT GAT	432
Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met Glu Asp Asp	
130 135 140	
CTG GTG ATT TCA TTT CAG TTA ATG CTA TGT GTC CTT GAC TAT TTT ATT	480
Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val Leu Asp Tyr Phe Ile	
145 150 155	
AAA CTC TCA CCT CCC ATG TTG CTC AAA GAA CCA TAT AAA ACA GCT GTT	528
Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys Thr Ala Val	
160 165 170	
ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG CGA GGT CAG AAC AGG	576
Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg	
175 180 185 190	
AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT ACA AGA ATT ATT GAA	624
Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu	
195 200 205	
GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT GAG GTG AAA AAT GTT	672
Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val	
210 215 220	
TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT CTT GGA CTT GTA ACA	720
Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr	
225 230 235	
TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT AAA CGA TAC GAA GAA	768
Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu	
240 245 250	
ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA TTA TTT TTG GAT CAT	816
Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His	
255 260 265 270	
GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT TTT GAA ACA CAG AGA	864
Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg	
275 280 285	
ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG AAT GTA ATT CCT CCA	912
Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro	
290 295 300	

CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC CAA CAA TTA ATG ATG His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met 305 310 315	960
ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA AAT CTG ATT TCC TAT Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr 320 325 330	1008
TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT ATA CTG AAA AGA GTG Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val 335 340 345 350	1056
AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT GCT AAA GCT GTG GGA Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly 355 360 365	1104
CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC AAA CTT GGA GTT CGC Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg 370 375 380	1152
TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA TCA GAA GAA GAA CGA Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg 385 390 395	1200
TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT GAC AAC ATT TTT CAT Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His 400 405 410	1248
ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA ATG GCC ACA TAT AGC Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser 415 420 425 430	1296
AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA GAT TTG TCT TTC CCA Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro 435 440 445	1344
TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT GAT TTT TAC AAA GTG Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val 450 455 460	1392
ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG ACA AGA GAA ATG ATA Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile 465 470 475	1440
AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG GAA TCC CTT GCA TGG Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp 480 485 490	1488
CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA CAA TCA AAG GAC CGA Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg 495 500 505 510	1536

GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT CCT CTT AAT CTT CCT Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro 515 520 525	1584
CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT CTT TCT CCT GTA AGA Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg 530 535 540	1632
TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT GCA AAT Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn 545 550 555	1680
GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA TTG AAA Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys 560 565 570	1728
TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA GCC TAT Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr 575 580 585 590	1776
CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG TCT GAG CAC CCA GAA Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His Pro Glu 595 600 605	1824
TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC ACC CTG CAG AAT GAG Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln Asn Glu 610 615 620	1872
TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA ATT ATG ATG TGT TCC Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met Met Cys Ser 625 630 635	1920
ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC CTT AAA TTC AAA ATC Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys Phe Lys Ile 640 645 650	1968
ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT GTT CAG GAG ACA TTC Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln Glu Thr Phe 655 660 665 670	2016
AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT TCT ATT ATA GTA TTC Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe 675 680 685	2064
TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA AAT ATT TTG CAG TAT Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr 690 695 700	2112
GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA CCT CAC ATT CCT CGA Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His Ile Pro Arg 705 710 715	2160

AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG ATT CCT GGA GGG AAC Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn 720 725 730	2208
ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA ATT TCA GAA GGT CTG Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu 735 740 745 750	2256
CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA ATC TTA GTA TCA ATT Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val Ser Ile 755 760 765	2304
GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG AAA ATA AAT CAG ATG Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn Gln Met 770 775 780	2352
GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT GCT GAA GGA AGC AAC Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly Ser Asn 785 790 795	2400
CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT ATT GAA GGA TCA GAT Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp 800 805 810	2448
GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG TCC AAA TTT CAG CAG Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln 815 820 825 830	2496
AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA ATG CAA AAG CAG AAA Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys 835 840 845	2544
ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG AAA TGAGGATCTC Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys 850 855	2593
AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG TCTCTCACAG ATGTGACTGT	2653
ATAACTTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT TTTGTGGATA	2713
TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG TTAGTCATTG	2773
TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA TTTAAAAAGT TGTAGCAGAT	2833
TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG GCCCTAGAGT	2893
GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG CATATAGGTG	2953
ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC ATGAACACCC	3013
TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTTGA TTGACTGCCC ATTCACCAAA	3073
ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA AATTACTAAT	3133

Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg Ser Ala
180 185 190

Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu Val Leu
195 200 205

Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val Tyr Phe
210 215 220

Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr Ser Asn
225 230 235 240

Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile Tyr
245 250 255

Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp Lys
260 265 270

Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro
275 280 285

Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr
290 295 300

Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met Ile Leu
305 310 315 320

Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn
325 330 335

Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys Asp
340 345 350

Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln Gly
355 360 365

Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr
370 375 380

Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu Ser
385 390 395 400

Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met Ser
405 410 415

Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg Ser
420 425 430

Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile
435 440 445

Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu
450 455 460

Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	Glu	Met	Ile	Lys	His	
465					470					475					480	
Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	Leu	Ala	Trp	Leu	Ser	
				485					490					495		
Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	Lys	Asp	Arg	Glu	Gly	
			500					505					510			
Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	Asn	Leu	Pro	Leu	Gln	
		515					520					525				
Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	Pro	Val	Arg	Ser	Pro	
	530					535						540				
Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	Thr	Ala	Asn	Ala	Glu	
545					550					555					560	
Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	Pro	Leu	Lys	Ser	Thr	
				565					570					575		
Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	Leu	Ala	Tyr	Leu	Arg	
			580					585					590			
Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	His	Pro	Glu	Leu	Glu	
		595					600					605				
His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	Gln	Asn	Glu	Tyr	Glu	
	610					615					620					
Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	Met	Cys	Ser	Met	Tyr	
625					630					635					640	
Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	Phe	Lys	Ile	Ile	Val	
				645					650					655		
Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	Glu	Thr	Phe	Lys	Arg	
			660					665					670			
Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	Ile	Val	Phe	Tyr	Asn	
	675						680					685				
Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	Leu	Gln	Tyr	Ala	Ser	
	690					695					700					
Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	Ile	Pro	Arg	Ser	Pro	
705					710					715					720	
Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	Gly	Gly	Asn	Ile	Tyr	
				725					730					735		
Ile	Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	Glu	Gly	Leu	Pro	Thr	
			740					745					750			

Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val Ser Ile Gly Glu
755 760 765

Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn Gln Met Val Cys
770 775 780

Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro
785 790 795 800

Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala
805 810 815

Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu
820 825 830

Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys Met Asn
835 840 845

Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys
850 855

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3161 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 7..2397

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	

AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG CGA GGT CAG	384
Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln	
115 120 125	
AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT ACA AGA ATT	432
Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile	
130 135 140	
ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT GAG GTG AAA	480
Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys	
145 150 155	
AAT GTT TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT CTT GGA CTT	528
Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu	
160 165 170	
GTA ACA TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT AAA CGA TAC	576
Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr	
175 180 185 190	
GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA TTA TTT TTG	624
Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu	
195 200 205	
GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT TTT GAA ACA	672
Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr	
210 215 220	
CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG AAT GTA ATT	720
Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile	
225 230 235	
CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC CAA CAA TTA	768
Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu	
240 245 250	
ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA AAT CTG ATT	816
Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile	
255 260 265 270	

TCC	TAT	TTT	AAC	AAC	TGC	ACA	GTG	AAT	CCA	AAA	GAA	AGT	ATA	CTG	AAA	864
Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	Lys	Glu	Ser	Ile	Leu	Lys	
			275						280					285		
AGA	GTG	AAG	GAT	ATA	GGA	TAC	ATC	TTT	AAA	GAG	AAA	TTT	GCT	AAA	GCT	912
Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	Glu	Lys	Phe	Ala	Lys	Ala	
			290					295					300			
GTG	GGA	CAG	GGT	TGT	GTC	GAA	ATT	GGA	TCA	CAG	CGA	TAC	AAA	CTT	GGA	960
Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	Gln	Arg	Tyr	Lys	Leu	Gly	
		305					310					315				
GTT	CGC	TTG	TAT	TAC	CGA	GTA	ATG	GAA	TCC	ATG	CTT	AAA	TCA	GAA	GAA	1008
Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	Met	Leu	Lys	Ser	Glu	Glu	
	320					325					330					
GAA	CGA	TTA	TCC	ATT	CAA	AAT	TTT	AGC	AAA	CTT	CTG	AAT	GAC	AAC	ATT	1056
Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	Leu	Leu	Asn	Asp	Asn	Ile	
335					340					345					350	
TTT	CAT	ATG	TCT	TTA	TTG	GCG	TGC	GCT	CTT	GAG	GTT	GTA	ATG	GCC	ACA	1104
Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	Glu	Val	Val	Met	Ala	Thr	
			355						360					365		
TAT	AGC	AGA	AGT	ACA	TCT	CAG	AAT	CTT	GAT	TCT	GGA	ACA	GAT	TTG	TCT	1152
Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	Ser	Gly	Thr	Asp	Leu	Ser	
			370					375					380			
TTC	CCA	TGG	ATT	CTG	AAT	GTG	CTT	AAT	TTA	AAA	GCC	TTT	GAT	TTT	TAC	1200
Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	Lys	Ala	Phe	Asp	Phe	Tyr	
		385					390					395				
AAA	GTG	ATC	GAA	AGT	TTT	ATC	AAA	GCA	GAA	GGC	AAC	TTG	ACA	AGA	GAA	1248
Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	Gly	Asn	Leu	Thr	Arg	Glu	
	400					405					410					
ATG	ATA	AAA	CAT	TTA	GAA	CGA	TGT	GAA	CAT	CGA	ATC	ATG	GAA	TCC	CTT	1296
Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	Arg	Ile	Met	Glu	Ser	Leu	
415					420					425					430	
GCA	TGG	CTC	TCA	GAT	TCA	CCT	TTA	TTT	GAT	CTT	ATT	AAA	CAA	TCA	AAG	1344
Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	Leu	Ile	Lys	Gln	Ser	Lys	
				435					440					445		
GAC	CGA	GAA	GGA	CCA	ACT	GAT	CAC	CTT	GAA	TCT	GCT	TGT	CCT	CTT	AAT	1392
Asp	Arg	Glu	Gly	Pro	Thr	Asp	His	Leu	Glu	Ser	Ala	Cys	Pro	Leu	Asn	
			450					455					460			
CTT	CCT	CTC	CAG	AAT	AAT	CAC	ACT	GCA	GCA	GAT	ATG	TAT	CTT	TCT	CCT	1440
Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	Pro	
		465														

GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT	1488
Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr	
480 485 490	
GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA	1536
Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro	
495 500 505 510	
TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA	1584
Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu	
515 520 525	
GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG TCT GAG CAC	1632
Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His	
530 535 540	
CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC ACC CTG CAG	1680
Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln	
545 550 555	
AAT GAG TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA ATT ATG ATG	1728
Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met Met	
560 565 570	
TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC CTT AAA TTC	1776
Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys Phe	
575 580 585 590	
AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT GTT CAG GAG	1824
Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln Glu	
595 600 605	
ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT TCT ATT ATA	1872
Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile	
610 615 620	
GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA AAT ATT TTG	1920
Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile Leu	
625 630 635	
CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA CCT CAC ATT	1968
Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His Ile	
640 645 650	
CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG ATT CCT GGA	2016
Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly	
655 660 665 670	
GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA ATT TCA GAA	2064
Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu	
675 680 685	

GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA ATC TTA GTA	2112
Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val	
690 695 700	
TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG AAA ATA AAT	2160
Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn	
705 710 715	
CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT GCT GAA GGA	2208
Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly	
720 725 730	
AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT ATT GAA GGA	2256
Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly	
735 740 745 750	
TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG TCC AAA TTT	2304
Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe	
755 760 765	
CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA ATG CAA AAG	2352
Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys	
770 775 780	
CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA GAG AAA	2397
Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys	
785 790 795	
TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG TCTCTCACAG	2457
ATGTGACTGT ATAACCTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC TTCAGCTCTT	2517
TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC ACTTGAAATG	2577
TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA TTTAAAAAGT	2637
TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT AGTAAGAATG	2697
GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT TCTTTTGTAG	2757
CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT TTAATTTAAC	2817
ATGAACACCC TTAGAAAATG TGTCTATCT ATCTTCCAAA TGCAATTTGA TTGACTGCCC	2877
ATTACACAAA ATTATCCTGA ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA	2937
AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT ACTGTGTGCT	2997
TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA	3057
TACTATCATA CTACTGAAAC AGATTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT	3117
TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT	3161

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 797 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala Ala Ala
1 5 10 15
Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu Glu Asp
20 25 30
Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg Leu Glu
35 40 45
Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln Lys Leu
50 55 60
Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp Glu Lys
65 70 75 80
Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys Lys Lys
85 90 95
Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Ala Val
100 105 110
Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg
115 120 125
Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu
130 135 140
Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp Glu Val Lys Asn Val
145 150 155 160
Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser Leu Gly Leu Val Thr
165 170 175
Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu
180 185 190
Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe Leu Asp His
195 200 205
Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg
210 215 220

Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro
225 230 235 240

His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met
245 250 255

Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr
260 265 270

Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val
275 280 285

Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly
290 295 300

Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg
305 310 315 320

Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg
325 330 335

Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn Ile Phe His
340 345 350

Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser
355 360 365

Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro
370 375 380

Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val
385 390 395 400

Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile
405 410 415

Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp
420 425 430

Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg
435 440 445

Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro
450 455 460

Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg
465 470 475 480

Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn
485 490 495

Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys
500 505 510

Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr	515	520	525
Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu Ser Glu His Pro Glu	530	535	540
Leu Glu His Ile Ile Trp Thr Leu Phe Gln His Thr Leu Gln Asn Glu	545	550	555 560
Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln Ile Met Met Cys Ser	565	570	575
Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp Leu Lys Phe Lys Ile	580	585	590
Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala Val Gln Glu Thr Phe	595	600	605
Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe	610	615	620
Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr	625	630	635 640
Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile Pro His Ile Pro Arg	645	650	655
Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn	660	665	670
Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu	675	680	685
Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg Ile Leu Val Ser Ile	690	695	700
Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln Lys Ile Asn Gln Met	705	710	715 720
Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser Ala Glu Gly Ser Asn	725	730	735
Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp	740	745	750
Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu Ser Lys Phe Gln Gln	755	760	765
Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg Met Gln Lys Gln Lys	770	775	780
Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu Glu Lys	785	790	795

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3377 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2613

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	
CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GAT GAG ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	384
Asp Glu Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
115 120 125	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	432
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
130 135 140	
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG	480
Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu	
145 150 155	

TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu 160 165 170	528
ACA CAA CCC AGC AGT TCG ATG GTC GCT GTT ATA CCC ATT AAT GGT TCA Thr Gln Pro Ser Ser Met Val Ala Val Ile Pro Ile Asn Gly Ser 175 180 185 190	576
CCT CGA ACA CCC AGG CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA Pro Arg Thr Pro Arg Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys 195 200 205	624
CAA CTA GAA AAT GAT ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT Gln Leu Glu Asn Asp Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His 210 215 220	672
GAA TGT AAT ATA GAT GAG GTG AAA AAT GTT TAT TTC AAA AAT TTT ATA Glu Cys Asn Ile Asp Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile 225 230 235	720
CCT TTT ATG AAT TCT CTT GGA CTT GTA ACA TCT AAT GGA CTT CCA GAG Pro Phe Met Asn Ser Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu 240 245 250	768
GTT GAA AAT CTT TCT AAA CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys 255 260 265 270	816
GAT CTA GAT GCA AGA TTA TTT TTG GAT CAT GAT AAA ACT CTT CAG ACT Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr 275 280 285	864
GAT TCT ATA GAC AGT TTT GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn 290 295 300	912
CTT GAT GAA GAG GTG AAT GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr Pro Val Arg Thr 305 310 315	960
GTT ATG AAC ACT ATC CAA CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT Val Met Asn Thr Ile Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser 320 325 330	1008
GAT CAA CCT TCA GAA AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val 335 340 345 350	1056
AAT CCA AAA GAA AGT ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile 355 360 365	1104

TTT AAA GAG AAA TTT GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT	1152
Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile	
370 375 380	
GGA TCA CAG CGA TAC AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG	1200
Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met	
385 390 395	
GAA TCC ATG CTT AAA TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT	1248
Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe	
400 405 410	
AGC AAA CTT CTG AAT GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC	1296
Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys	
415 420 425 430	
GCT CTT GAG GTT GTA ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT	1344
Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn	
435 440 445	
CTT GAT TCT GGA ACA GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT	1392
Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu	
450 455 460	
AAT TTA AAA GCC TTT GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA	1440
Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys	
465 470 475	
GCA GAA GGC AAC TTG ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT	1488
Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys	
480 485 490	
GAA CAT CGA ATC ATG GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA	1536
Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu	
495 500 505 510	
TTT GAT CTT ATT AAA CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC	1584
Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His	
515 520 525	
CTT GAA TCT GCT TGT CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT	1632
Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr	
530 535 540	
GCA GCA GAT ATG TAT CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT	1680
Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly	
545 550 555	
TCA ACT ACG CGT GTA AAT TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC	1728
Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr	
560 565 570	

TCA	GCC	TTC	CAG	ACC	CAG	AAG	CCA	TTG	AAA	TCT	ACC	TCT	CTT	TCA	CTG	1776
Ser	Ala	Phe	Gln	Thr	Gln	Lys	Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	
575					580					585					590	
TTT	TAT	AAA	AAA	GTG	TAT	CGG	CTA	GCC	TAT	CTC	CGG	CTA	AAT	ACA	CTT	1824
Phe	Tyr	Lys	Lys	Val	Tyr	Arg	Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	
				595					600					605		
TGT	GAA	CGC	CTT	CTG	TCT	GAG	CAC	CCA	GAA	TTA	GAA	CAT	ATC	ATC	TGG	1872
Cys	Glu	Arg	Leu	Leu	Ser	Glu	His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	
			610					615					620			
ACC	CTT	TTC	CAG	CAC	ACC	CTG	CAG	AAT	GAG	TAT	GAA	CTC	ATG	AGA	GAC	1920
Thr	Leu	Phe	Gln	His	Thr	Leu	Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	
		625					630						635			
AGG	CAT	TTG	GAC	CAA	ATT	ATG	ATG	TGT	TCC	ATG	TAT	GGC	ATA	TGC	AAA	1968
Arg	His	Leu	Asp	Gln	Ile	Met	Met	Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	
	640					645					650					
GTG	AAG	AAT	ATA	GAC	CTT	AAA	TTC	AAA	ATC	ATT	GTA	ACA	GCA	TAC	AAG	2016
Val	Lys	Asn	Ile	Asp	Leu	Lys	Phe	Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	
655					660					665					670	
GAT	CTT	CCT	CAT	GCT	GTT	CAG	GAG	ACA	TTC	AAA	CGT	GTT	TTG	ATC	AAA	2064
Asp	Leu	Pro	His	Ala	Val	Gln	Glu	Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	
				675					680					685		
GAA	GAG	GAG	TAT	GAT	TCT	ATT	ATA	GTA	TTC	TAT	AAC	TCG	GTC	TTC	ATG	2112
Glu	Glu	Glu	Tyr	Asp	Ser	Ile	Ile	Val	Phe	Tyr	Asn	Ser	Val	Phe	Met	
			690					695					700			
CAG	AGA	CTG	AAA	ACA	AAT	ATT	TTG	CAG	TAT	GCT	TCC	ACC	AGG	CCC	CCT	2160
Gln	Arg	Leu	Lys	Thr	Asn	Ile	Leu	Gln	Tyr	Ala	Ser	Thr	Arg	Pro	Pro	
		705					710					715				
ACC	TTG	TCA	CCA	ATA	CCT	CAC	ATT	CCT	CGA	AGC	CCT	TAC	AAG	TTT	CCT	2208
Thr	Leu	Ser	Pro	Ile	Pro	His	Ile	Pro	Arg	Ser	Pro	Tyr	Lys	Phe	Pro	
		720				725					730					
AGT	TCA	CCC	TTA	CGG	ATT	CCT	GGA	GGG	AAC	ATC	TAT	ATT	TCA	CCC	CTG	2256
Ser	Ser	Pro	Leu	Arg	Ile	Pro	Gly	Gly	Asn	Ile	Tyr	Ile	Ser	Pro	Leu	
735					740					745					750	
AAG	AGT	CCA	TAT	AAA	ATT	TCA	GAA	GGT	CTG	CCA	ACA	CCA	ACA	AAA	ATG	2304
Lys	Ser	Pro	Tyr	Lys	Ile	Ser	Glu	Gly	Leu	Pro	Thr	Pro	Thr	Lys	Met	
				755				760						765		
ACT	CCA	AGA	TCA	AGA	ATC	TTA	GTA	TCA	ATT	GGT	GAA	TCA	TTC	GGG	ACT	2352
Thr	Pro	Arg	Ser	Arg	Ile	Leu	Val	Ser	Ile	Gly	Glu	Ser	Phe	Gly	Thr	
			770					775					780			

TCT GAG AAG TTC CAG AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT Ser Glu Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg 785 790 795	2400
GTG CTC AAA AGA AGT GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA Val Leu Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys 800 805 810	2448
AAA CTA CGC TTT GAT ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys 815 820 825 830	2496
CAT CTC CCA GGA GAG TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT His Leu Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr 835 840 845	2544
TCT ACT CGA ACA CGA ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT Ser Thr Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp 850 855 860	2592
ACC TCA AAC AAG GAA GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT Thr Ser Asn Lys Glu Glu Lys 865	2643
GTACACCTCT GGATTCATTG TCTCTCACAG ATGTGACTGT ATAACCTTCC CAGGTTCTGT	2703
TTATGGCCAC ATTTAATATC TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT	2763
GTTTGGGTGA TTCCTAAGCC ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA	2823
ATCTTGTGTA AATCCTGCCA TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA	2883
AATTGCTGTG CTTTATGGAT AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG	2943
CCTGTCTGAC TACTTTGCCT TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT	3003
TAATTTATAT GTATATTTTT TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCCTATCT	3063
ATCTTCCAAA TGCAATTTGA TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC	3123
AAAAATGGAT ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT	3183
TTACTATTGG AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT	3243
AAAAGCTGGA AGCAAAGTAT AACCATATGA TACTATCATA CTACTGAAAC AGATTTTCATA	3303
CCTCAGAATG TAAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA	3363
GGATTATTGA TAGT	3377

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 869 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala Ala Ala
1 5 10 15
Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu Glu Asp
20 25 30
Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg Leu Glu
35 40 45
Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln Lys Leu
50 55 60
Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp Glu Lys
65 70 75 80
Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys Lys Lys
85 90 95
Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu Asp Glu
100 105 110
Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile Ser Val
115 120 125
His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr Lys Val
130 135 140
Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu Phe Ala
145 150 155 160
Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln
165 170 175
Pro Ser Ser Ser Met Val Ala Val Ile Pro Ile Asn Gly Ser Pro Arg
180 185 190
Thr Pro Arg Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu
195 200 205
Glu Asn Asp Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys
210 215 220

Asn	Ile	Asp	Glu	Val	Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	
225					230					235					240	
Met	Asn	Ser	Leu	Gly	Leu	Val	Thr	Ser	Asn	Gly	Leu	Pro	Glu	Val	Glu	
			245						250					255		
Asn	Leu	Ser	Lys	Arg	Tyr	Glu	Glu	Ile	Tyr	Leu	Lys	Asn	Lys	Asp	Leu	
			260					265						270		
Asp	Ala	Arg	Leu	Phe	Leu	Asp	His	Asp	Lys	Thr	Leu	Gln	Thr	Asp	Ser	
		275					280						285			
Ile	Asp	Ser	Phe	Glu	Thr	Gln	Arg	Thr	Pro	Arg	Lys	Ser	Asn	Leu	Asp	
	290					295					300					
Glu	Glu	Val	Asn	Val	Ile	Pro	Pro	His	Thr	Pro	Val	Arg	Thr	Val	Met	
305					310					315					320	
Asn	Thr	Ile	Gln	Gln	Leu	Met	Met	Ile	Leu	Asn	Ser	Ala	Ser	Asp	Gln	
			325						330						335	
Pro	Ser	Glu	Asn	Leu	Ile	Ser	Tyr	Phe	Asn	Asn	Cys	Thr	Val	Asn	Pro	
			340					345						350		
Lys	Glu	Ser	Ile	Leu	Lys	Arg	Val	Lys	Asp	Ile	Gly	Tyr	Ile	Phe	Lys	
		355					360					365				
Glu	Lys	Phe	Ala	Lys	Ala	Val	Gly	Gln	Gly	Cys	Val	Glu	Ile	Gly	Ser	
	370					375					380					
Gln	Arg	Tyr	Lys	Leu	Gly	Val	Arg	Leu	Tyr	Tyr	Arg	Val	Met	Glu	Ser	
385					390					395					400	
Met	Leu	Lys	Ser	Glu	Glu	Glu	Arg	Leu	Ser	Ile	Gln	Asn	Phe	Ser	Lys	
			405						410						415	
Leu	Leu	Asn	Asp	Asn	Ile	Phe	His	Met	Ser	Leu	Leu	Ala	Cys	Ala	Leu	
			420					425						430		
Glu	Val	Val	Met	Ala	Thr	Tyr	Ser	Arg	Ser	Thr	Ser	Gln	Asn	Leu	Asp	
		435					440					445				
Ser	Gly	Thr	Asp	Leu	Ser	Phe	Pro	Trp	Ile	Leu	Asn	Val	Leu	Asn	Leu	
	450					455					460					
Lys	Ala	Phe	Asp	Phe	Tyr	Lys	Val	Ile	Glu	Ser	Phe	Ile	Lys	Ala	Glu	
465					470					475					480	
Gly	Asn	Leu	Thr	Arg	Glu	Met	Ile	Lys	His	Leu	Glu	Arg	Cys	Glu	His	
			485					490						495		
Arg	Ile	Met	Glu	Ser	Leu	Ala	Trp	Leu	Ser	Asp	Ser	Pro	Leu	Phe	Asp	
		500						505						510		

Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu
515 520 525

Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala
530 535 540

Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr
545 550 555 560

Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala
565 570 575

Phe Gln Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr
580 585 590

Lys Lys Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu
595 600 605

Arg Leu Leu Ser Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu
610 615 620

Phe Gln His Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His
625 630 635 640

Leu Asp Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys
645 650 655

Asn Ile Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu
660 665 670

Pro His Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu
675 680 685

Glu Tyr Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg
690 695 700

Leu Lys Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu
705 710 715 720

Ser Pro Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser
725 730 735

Pro Leu Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser
740 745 750

Pro Tyr Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro
755 760 765

Arg Ser Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu
770 775 780

Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu
785 790 795 800

Lys	Arg	Ser	Ala	Glu	Gly	Ser	Asn	Pro	Pro	Lys	Pro	Leu	Lys	Lys	Leu
				805					810					815	
Arg	Phe	Asp	Ile	Glu	Gly	Ser	Asp	Glu	Ala	Asp	Gly	Ser	Lys	His	Leu
			820					825					830		
Pro	Gly	Glu	Ser	Lys	Phe	Gln	Gln	Lys	Leu	Ala	Glu	Met	Thr	Ser	Thr
			835				840					845			
Arg	Thr	Arg	Met	Gln	Lys	Gln	Lys	Met	Asn	Asp	Ser	Met	Asp	Thr	Ser
	850					855					860				
Asn	Lys	Glu	Glu	Lys											
865															

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3383 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2619

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CGCGTC	ATG	CCG	CCC	AAA	ACC	CCC	CGA	AAA	ACG	GCC	GCC	ACC	GCC	GCC		48
Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala			
1				5					10							
GCT	GCC	GCC	GCG	GAA	CCC	CCG	GCA	CCG	CCG	CCG	CCG	CCC	CCT	CCT	GAG	96
Ala	Ala	Ala	Ala	Glu	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	
15				20					25						30	
GAG	GAC	CCA	GAG	CAG	GAC	AGC	GGC	CCG	GAG	GAC	CTG	CCT	CTC	GTC	AGG	144
Glu	Asp	Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	
				35				40						45		
CTT	GAG	TTT	GAA	GAA	ACA	GAA	GAA	CCT	GAT	TTT	ACT	GCA	TTA	TGT	CAG	192
Leu	Glu	Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	
			50					55					60			
AAA	TTA	AAG	ATA	CCA	GAT	CAT	GTC	AGA	GAG	AGA	GCT	TGG	TTA	ACT	TGG	240
Lys	Leu	Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	
		65					70				75					
GAG	AAA	GTT	TCA	TCT	GTG	GAT	GGA	GTA	TTG	GGA	GGT	TAT	ATT	CAA	AAG	288
Glu	Lys	Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	
	80					85					90					

AAA Lys 95	AAG Lys	GAA Glu	CTG Leu	TGG Trp	GGA Gly 100	ATC Ile	TGT Cys	ATC Ile	TTT Phe	ATT Ile 105	GCA Ala	GCA Ala	GTT Val	GAC Asp	CTA Leu 110	336
GAT Asp	GAG Glu	ATG Met	TCG Ser	TTC Phe 115	ACT Thr	TTT Phe	ACT Thr	GAG Glu	CTA Leu 120	CAG Gln	AAA Lys	AAC Asn	ATA Ile	GAA Glu 125	ATC Ile	384
AGT Ser	GTC Val	CAT His	AAA Lys 130	TTC Phe	TTT Phe	AAC Asn	TTA Leu	CTA Leu 135	AAA Lys	GAA Glu	ATT Ile	GAT Asp	ACC Thr 140	AGT Ser	ACC Thr	432
AAA Lys	GTT Val 145	GAT Asp	AAT Asn	GCT Ala	ATG Met	TCA Ser	AGA Arg 150	CTG Leu	TTG Leu	AAG Lys	AAG Lys	TAT Tyr 155	GAT Asp	GTA Val	TTG Leu	480
TTT Phe 160	GCA Ala	CTC Leu	TTC Phe	AGC Ser	AAA Lys 165	TTG Leu	GAA Glu	AGG Arg	ACA Thr	TGT Cys 170	GAA Glu	CTT Leu	ATA Ile	TAT Tyr	TTG Leu	528
ACA Thr 175	CAA Gln	CCC Pro	AGC Ser	AGT Ser	TCG Ser 180	ATA Ile	TCT Ser	ACT Thr	GAA Glu 185	ATA Ile	AAT Asn	TCT Ser	GCA Ala	TTG Leu 190	GTG Val	576
CTA Leu	AAA Lys	GTT Val	TCT Ser 195	TGG Trp	ATC Ile	ACA Thr	TTT Phe	TTA Leu 200	GCT Ala	AAA Lys	GGG Gly	GAA Glu 205	GTA Val	TTA Leu		624
CAA Gln	ATG Met	GAA Glu 210	GAT Asp	GAT Asp	CTG Leu	GTG Val	ATT Ile 215	TCA Ser	TTT Phe	CAG Gln	TTA Leu	ATG Met 220	CTA Leu	TGT Cys	GTC Val	672
CTT Leu	GAC Asp 225	TAT Tyr	TTT Phe	ATT Ile	AAA Lys	CTC Leu	TCA Ser 230	CCT Pro	CCC Pro	ATG Met	TTG Leu 235	CTC Leu	AAA Lys	GAA Glu	CCA Pro	720
TAT Tyr 240	AAA Lys	ACA Thr	GGG Gly	TCG Ser	AAT Asn	TCT Ser 245	CTT Leu	GGA Gly	CTT Leu	GTA Val 250	ACA Thr	TCT Ser	AAT Asn	GGA Gly	CTT Leu	768
CCA Pro 255	GAG Glu	GTT Val	GAA Glu	AAT Asn 260	CTT Leu	TCT Ser	AAA Lys	CGA Arg	TAC Tyr	GAA Glu 265	GAA Glu	ATT Ile	TAT Tyr	CTT Leu 270	AAA Lys	816
AAT Asn	AAA Lys	GAT Asp	CTA Leu 275	GAT Asp	GCA Ala	AGA Arg	TTA Leu	TTT Phe 280	TTG Leu	GAT Asp	CAT His	GAT Asp	AAA Lys 285	ACT Thr	CTT Leu	864
CAG Gln	ACT Thr	GAT Asp	TCT Ser 290	ATA Ile	GAC Asp	AGT Ser	TTT Phe 295	GAA Glu	ACA Thr	CAG Gln	AGA Arg	ACA Thr 300	CCA Pro	CGA Arg	AAA Lys	912

AGT AAC CTT GAT GAA GAG GTG AAT GTA ATT CCT CCA CAC ACT CCA GTT	960
Ser Asn Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr Pro Val	
305 310 315	
AGG ACT GTT ATG AAC ACT ATC CAA CAA TTA ATG ATG ATT TTA AAT TCA	1008
Arg Thr Val Met Asn Thr Ile Gln Gln Leu Met Met Ile Leu Asn Ser	
320 325 330	
GCA AGT GAT CAA CCT TCA GAA AAT CTG ATT TCC TAT TTT AAC AAC TGC	1056
Ala Ser Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys	
335 340 345 350	
ACA GTG AAT CCA AAA GAA AGT ATA CTG AAA AGA GTG AAG GAT ATA GGA	1104
Thr Val Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile Gly	
355 360 365	
TAC ATC TTT AAA GAG AAA TTT GCT AAA GCT GTG GGA CAG GGT TGT GTC	1152
Tyr Ile Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys Val	
370 375 380	
GAA ATT GGA TCA CAG CGA TAC AAA CTT GGA GTT CGC TTG TAT TAC CGA	1200
Glu Ile Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg	
385 390 395	
GTA ATG GAA TCC ATG CTT AAA TCA GAA GAA GAA CGA TTA TCC ATT CAA	1248
Val Met Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln	
400 405 410	
AAT TTT AGC AAA CTT CTG AAT GAC AAC ATT TTT CAT ATG TCT TTA TTG	1296
Asn Phe Ser Lys Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu	
415 420 425 430	
GCG TGC GCT CTT GAG GTT GTA ATG GCC ACA TAT AGC AGA AGT ACA TCT	1344
Ala Cys Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser	
435 440 445	
CAG AAT CTT GAT TCT GGA ACA GAT TTG TCT TTC CCA TGG ATT CTG AAT	1392
Gln Asn Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn	
450 455 460	
GTG CTT AAT TTA AAA GCC TTT GAT TTT TAC AAA GTG ATC GAA AGT TTT	1440
Val Leu Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe	
465 470 475	
ATC AAA GCA GAA GGC AAC TTG ACA AGA GAA ATG ATA AAA CAT TTA GAA	1488
Ile Lys Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu Glu	
480 485 490	
CGA TGT GAA CAT CGA ATC ATG GAA TCC CTT GCA TGG CTC TCA GAT TCA	1536
Arg Cys Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp Ser	
495 500 505 510	

1000
 900
 800
 700
 600
 500
 400
 300
 200
 100
 0

CCT TTA TTT GAT CTT ATT AAA CAA TCA AAG GAC CGA GAA GGA CCA ACT	1584
Pro Leu Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr	
515 520 525	
GAT CAC CTT GAA TCT GCT TGT CCT CTT AAT CTT CCT CTC CAG AAT AAT	1632
Asp His Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn	
530 535 540	
CAC ACT GCA GCA GAT ATG TAT CTT TCT CCT GTA AGA TCT CCA AAG AAA	1680
His Thr Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys	
545 550 555	
AAA GGT TCA ACT ACG CGT GTA AAT TCT ACT GCA AAT GCA GAG ACA CAA	1728
Lys Gly Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr Gln	
560 565 570	
GCA ACC TCA GCC TTC CAG ACC CAG AAG CCA TTG AAA TCT ACC TCT CTT	1776
Ala Thr Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu	
575 580 585 590	
TCA CTG TTT TAT AAA AAA GTG TAT CGG CTA GCC TAT CTC CGG CTA AAT	1824
Ser Leu Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn	
595 600 605	
ACA CTT TGT GAA CGC CTT CTG TCT GAG CAC CCA GAA TTA GAA CAT ATC	1872
Thr Leu Cys Glu Arg Leu Leu Ser Glu His Pro Glu Leu Glu His Ile	
610 615 620	
ATC TGG ACC CTT TTC CAG CAC ACC CTG CAG AAT GAG TAT GAA CTC ATG	1920
Ile Trp Thr Leu Phe Gln His Thr Leu Gln Asn Glu Tyr Glu Leu Met	
625 630 635	
AGA GAC AGG CAT TTG GAC CAA ATT ATG ATG TGT TCC ATG TAT GGC ATA	1968
Arg Asp Arg His Leu Asp Gln Ile Met Met Cys Ser Met Tyr Gly Ile	
640 645 650	
TGC AAA GTG AAG AAT ATA GAC CTT AAA TTC AAA ATC ATT GTA ACA GCA	2016
Cys Lys Val Lys Asn Ile Asp Leu Lys Phe Lys Ile Ile Val Thr Ala	
655 660 665 670	
TAC AAG GAT CTT CCT CAT GCT GTT CAG GAG ACA TTC AAA CGT GTT TTG	2064
Tyr Lys Asp Leu Pro His Ala Val Gln Glu Thr Phe Lys Arg Val Leu	
675 680 685	
ATC AAA GAA GAG GAG TAT GAT TCT ATT ATA GTA TTC TAT AAC TCG GTC	2112
Ile Lys Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe Tyr Asn Ser Val	
690 695 700	
TTC ATG CAG AGA CTG AAA ACA AAT ATT TTG CAG TAT GCT TCC ACC AGG	2160
Phe Met Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg	
705 710 715	

CCC Pro	CCT Pro	ACC Thr	TTG Leu	TCA Ser	CCA Pro	ATA Ile	CCT Pro	CAC His	ATT Ile	CCT Pro	CGA Arg	AGC Ser	CCT Pro	TAC Tyr	AAG Lys	2208	
720				725				730									
TTT Phe	CCT Pro	AGT Ser	TCA Ser	CCC Pro	TTA Leu	CGG Arg	ATT Ile	CCT Pro	GGA Gly	GGG Gly	AAC Asn	ATC Ile	TAT Tyr	ATT Ile	TCA Ser	2256	
735				740				745				750					
CCC Pro	CTG Leu	AAG Lys	AGT Ser	CCA Pro	TAT Tyr	AAA Lys	ATT Ile	TCA Ser	GAA Glu	GGT Gly	CTG Leu	CCA Pro	ACA Thr	CCA Pro	ACA Thr	2304	
755				760				765									
AAA Lys	ATG Met	ACT Thr	CCA Pro	AGA Arg	TCA Ser	AGA Arg	ATC Ile	TTA Leu	GTA Val	TCA Ser	ATT Ile	GGT Gly	GAA Glu	TCA Ser	TTC Phe	2352	
770				775				780									
GGG Gly	ACT Thr	TCT Ser	GAG Glu	AAG Lys	TTC Phe	CAG Gln	AAA Lys	ATA Ile	AAT Asn	CAG Gln	ATG Met	GTA Val	TGT Cys	AAC Asn	AGC Ser	2400	
785				790				795									
GAC Asp	CGT Arg	GTG Val	CTC Leu	AAA Lys	AGA Arg	AGT Ser	GCT Ala	GAA Glu	GGA Gly	AGC Ser	AAC Asn	CCT Pro	CCT Pro	AAA Lys	CCA Pro	2448	
800				805				810									
CTG Leu	AAA Lys	AAA Lys	CTA Leu	CGC Arg	TTT Phe	GAT Asp	ATT Ile	GAA Glu	GGA Gly	TCA Ser	GAT Asp	GAA Glu	GCA Ala	GAT Asp	GGA Gly	2496	
815				820				825				830					
AGT Ser	AAA Lys	CAT His	CTC Leu	CCA Pro	GGA Gly	GAG Glu	TCC Ser	AAA Lys	TTT Phe	CAG Gln	CAG Gln	AAA Lys	CTG Leu	GCA Ala	GAA Glu	2544	
835				840				845									
ATG Met	ACT Thr	TCT Ser	ACT Thr	CGA Arg	ACA Thr	CGA Arg	ATG Met	CAA Gln	AAG Lys	CAG Gln	AAA Lys	ATG Met	AAT Asn	GAT Asp	AGC Ser	2592	
850				855				860									
ATG Met	GAT Asp	ACC Thr	TCA Ser	AAC Asn	AAG Lys	GAA Glu	GAG Glu	AAA Lys	TGAGGATCTC				AGGACCTTGG			2639	
865				870													
TGGACACTGT			GTACACCTCT			GGATTTCATTG			TCTCTCACAG			ATGTGACTGT			ATAACTTTCC		2699
CAGGTTCTGT			TTATGGCCAC			ATTTAATATC			TTCAGCTCTT			TTTGTGGATA			TAAAATGTGC		2759
AGATGCAATT			GTTTGGGTGA			TTCCTAAGCC			ACTTGAAATG			TTAGTCATTG			TTATTTATAC		2819
AAGATTGAAA			ATCTTGTGTA			AATCCTGCCA			TTTAAAAAGT			TGTAGCAGAT			TGTTTCCTCT		2879
TCCAAAGTAA			AATTGCTGTG			CTTTATGGAT			AGTAAGAATG			GCCCTAGAGT			GGGAGTCCTG		2939
ATAACCCAGG			CCTGTCCTGAC			TACTTTGCCT			TCTTTTGTAG			CATATAGGTG			ATGTTTGCTC		2999
TTGTTTTTAT			TAATTTATAT			GTATATTTTT			TTAATTTAAC			ATGAACACCC			TTAGAAAAATG		3059

TGTCCTATCT ATCTTCCAAA TGCAATTTGA TTGACTGCCC ATTCACCAAAA ATTATCCTGA	3119
ACTCTTCTGC AAAAATGGAT ATTATTAGAA ATTAGAAAAA AATTACTAAT TTTACACATT	3179
AGATTTTATT TTAATTATTGG AATCTGATAT ACTGTGTGCT TGTTTTATAA AATTTTGCTT	3239
TTAATTAAAT AAAAGCTGGA AGCAAAGTAT AACCATATGA TACTATCATA CTAAGTAAAC	3299
AGATTTTCATA CCTCAGAATG TAAAAGAACT TACTGATTAT TTTCTTCATC CAACTTATGT	3359
TTTTTAAATGA GGATTATTGA TAGT	3383

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 871 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Met	Pro	Pro	Lys	Thr	Pro	Arg	Lys	Thr	Ala	Ala	Thr	Ala	Ala	Ala	Ala	1	5	10	15
Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp	20	25	30	
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	35	40	45	
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	50	55	60	
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	65	70	75	80
Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	85	90	95	
Glu	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Asp	Glu	100	105	110	
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	115	120	125	
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	130	135	140	
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	145	150	155	160

Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu Thr Gln
165 170 175

Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val Leu Lys
180 185 190

Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu Gln Met
195 200 205

Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val Leu Asp
210 215 220

Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro Tyr Lys
225 230 235 240

Thr Gly Ser Asn Ser Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu
245 250 255

Val Glu Asn Leu Ser Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys
260 265 270

Asp Leu Asp Ala Arg Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr
275 280 285

Asp Ser Ile Asp Ser Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn
290 295 300

Leu Asp Glu Glu Val Asn Val Ile Pro Pro His Thr Pro Val Arg Thr
305 310 315 320

Val Met Asn Thr Ile Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser
325 330 335

Asp Gln Pro Ser Glu Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val
340 345 350

Asn Pro Lys Glu Ser Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile
355 360 365

Phe Lys Glu Lys Phe Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile
370 375 380

Gly Ser Gln Arg Tyr Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met
385 390 395 400

Glu Ser Met Leu Lys Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe
405 410 415

Ser Lys Leu Leu Asn Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys
420 425 430

Ala Leu Glu Val Val Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn
435 440 445

Leu Asp Ser Gly Thr Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu
450 455 460

Asn Leu Lys Ala Phe Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys
465 470 475 480

Ala Glu Gly Asn Leu Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys
485 490 495

Glu His Arg Ile Met Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu
500 505 510

Phe Asp Leu Ile Lys Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His
515 520 525

Leu Glu Ser Ala Cys Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr
530 535 540

Ala Ala Asp Met Tyr Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly
545 550 555 560

Ser Thr Thr Arg Val Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr
565 570 575

Ser Ala Phe Gln Thr Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu
580 585 590

Phe Tyr Lys Lys Val Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu
595 600 605

Cys Glu Arg Leu Leu Ser Glu His Pro Glu Leu Glu His Ile Ile Trp
610 615 620

Thr Leu Phe Gln His Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp
625 630 635 640

Arg His Leu Asp Gln Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys
645 650 655

Val Lys Asn Ile Asp Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys
660 665 670

Asp Leu Pro His Ala Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys
675 680 685

Glu Glu Glu Tyr Asp Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met
690 695 700

Gln Arg Leu Lys Thr Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro
705 710 715 720

Thr Leu Ser Pro Ile Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro
725 730 735

Ser Ser Pro Leu Arg Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu
740 745 750

Lys Ser Pro Tyr Lys Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met
755 760 765

Thr Pro Arg Ser Arg Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr
770 775 780

Ser Glu Lys Phe Gln Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg
785 790 795 800

Val Leu Lys Arg Ser Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys
805 810 815

Lys Leu Arg Phe Asp Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys
820 825 830

His Leu Pro Gly Glu Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr
835 840 845

Ser Thr Arg Thr Arg Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp
850 855 860

Thr Ser Asn Lys Glu Glu Lys
865 870

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3554 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 7..2790

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CGCGTC ATG CCG CCC AAA ACC CCC CGA AAA ACG GCC GCC ACC GCC GCC	48
Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala	
1 5 10	
GCT GCC GCC GCG GAA CCC CCG GCA CCG CCG CCG CCC CCT CCT GAG	96
Ala Ala Ala Ala Glu Pro Pro Ala Pro Pro Pro Pro Pro Pro Glu	
15 20 25 30	
GAG GAC CCA GAG CAG GAC AGC GGC CCG GAG GAC CTG CCT CTC GTC AGG	144
Glu Asp Pro Glu Gln Asp Ser Gly Pro Glu Asp Leu Pro Leu Val Arg	
35 40 45	

CTT GAG TTT GAA GAA ACA GAA GAA CCT GAT TTT ACT GCA TTA TGT CAG	192
Leu Glu Phe Glu Glu Thr Glu Glu Pro Asp Phe Thr Ala Leu Cys Gln	
50 55 60	
AAA TTA AAG ATA CCA GAT CAT GTC AGA GAG AGA GCT TGG TTA ACT TGG	240
Lys Leu Lys Ile Pro Asp His Val Arg Glu Arg Ala Trp Leu Thr Trp	
65 70 75	
GAG AAA GTT TCA TCT GTG GAT GGA GTA TTG GGA GGT TAT ATT CAA AAG	288
Glu Lys Val Ser Ser Val Asp Gly Val Leu Gly Gly Tyr Ile Gln Lys	
80 85 90	
AAA AAG GAA CTG TGG GGA ATC TGT ATC TTT ATT GCA GCA GTT GAC CTA	336
Lys Lys Glu Leu Trp Gly Ile Cys Ile Phe Ile Ala Ala Val Asp Leu	
95 100 105 110	
GGT GAT ATG TCG TTC ACT TTT ACT GAG CTA CAG AAA AAC ATA GAA ATC	384
Gly Asp Met Ser Phe Thr Phe Thr Glu Leu Gln Lys Asn Ile Glu Ile	
115 120 125	
AGT GTC CAT AAA TTC TTT AAC TTA CTA AAA GAA ATT GAT ACC AGT ACC	432
Ser Val His Lys Phe Phe Asn Leu Leu Lys Glu Ile Asp Thr Ser Thr	
130 135 140	
AAA GTT GAT AAT GCT ATG TCA AGA CTG TTG AAG AAG TAT GAT GTA TTG	480
Lys Val Asp Asn Ala Met Ser Arg Leu Leu Lys Lys Tyr Asp Val Leu	
145 150 155	
TTT GCA CTC TTC AGC AAA TTG GAA AGG ACA TGT GAA CTT ATA TAT TTG	528
Phe Ala Leu Phe Ser Lys Leu Glu Arg Thr Cys Glu Leu Ile Tyr Leu	
160 165 170	
ACA CAA CCC AGC AGT TCG ATA TCT ACT GAA ATA AAT TCT GCA TTG GTG	576
Thr Gln Pro Ser Ser Ser Ile Ser Thr Glu Ile Asn Ser Ala Leu Val	
175 180 185 190	
CTA AAA GTT TCT TGG ATC ACA TTT TTA TTA GCT AAA GGG GAA GTA TTA	624
Leu Lys Val Ser Trp Ile Thr Phe Leu Leu Ala Lys Gly Glu Val Leu	
195 200 205	
CAA ATG GAA GAT GAT CTG GTG ATT TCA TTT CAG TTA ATG CTA TGT GTC	672
Gln Met Glu Asp Asp Leu Val Ile Ser Phe Gln Leu Met Leu Cys Val	
210 215 220	
CTT GAC TAT TTT ATT AAA CTC TCA CCT CCC ATG TTG CTC AAA GAA CCA	720
Leu Asp Tyr Phe Ile Lys Leu Ser Pro Pro Met Leu Leu Lys Glu Pro	
225 230 235	
TAT AAA ACA GCT GTT ATA CCC ATT AAT GGT TCA CCT CGA ACA CCC AGG	768
Tyr Lys Thr Ala Val Ile Pro Ile Asn Gly Ser Pro Arg Thr Pro Arg	
240 245 250	

CGA GGT CAG AAC AGG AGT GCA CGG ATA GCA AAA CAA CTA GAA AAT GAT	816
Arg Gly Gln Asn Arg Ser Ala Arg Ile Ala Lys Gln Leu Glu Asn Asp	
255 260 265 270	
ACA AGA ATT ATT GAA GTT CTC TGT AAA GAA CAT GAA TGT AAT ATA GAT	864
Thr Arg Ile Ile Glu Val Leu Cys Lys Glu His Glu Cys Asn Ile Asp	
275 280 285	
GAG GTG AAA AAT GTT TAT TTC AAA AAT TTT ATA CCT TTT ATG AAT TCT	912
Glu Val Lys Asn Val Tyr Phe Lys Asn Phe Ile Pro Phe Met Asn Ser	
290 295 300	
CTT GGA CTT GTA ACA TCT AAT GGA CTT CCA GAG GTT GAA AAT CTT TCT	960
Leu Gly Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser	
305 310 315	
AAA CGA TAC GAA GAA ATT TAT CTT AAA AAT AAA GAT CTA GAT GCA AGA	1008
Lys Arg Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg	
320 325 330	
TTA TTT TTG GAT CAT GAT AAA ACT CTT CAG ACT GAT TCT ATA GAC AGT	1056
Leu Phe Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser	
335 340 345 350	
TTT GAA ACA CAG AGA ACA CCA CGA AAA AGT AAC CTT GAT GAA GAG GTG	1104
Phe Glu Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val	
355 360 365	
AAT GTA ATT CCT CCA CAC ACT CCA GTT AGG ACT GTT ATG AAC ACT ATC	1152
Asn Val Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile	
370 375 380	
CAA CAA TTA ATG ATG ATT TTA AAT TCA GCA AGT GAT CAA CCT TCA GAA	1200
Gln Gln Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu	
385 390 395	
AAT CTG ATT TCC TAT TTT AAC AAC TGC ACA GTG AAT CCA AAA GAA AGT	1248
Asn Leu Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser	
400 405 410	
ATA CTG AAA AGA GTG AAG GAT ATA GGA TAC ATC TTT AAA GAG AAA TTT	1296
Ile Leu Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe	
415 420 425 430	
GCT AAA GCT GTG GGA CAG GGT TGT GTC GAA ATT GGA TCA CAG CGA TAC	1344
Ala Lys Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr	
435 440 445	
AAA CTT GGA GTT CGC TTG TAT TAC CGA GTA ATG GAA TCC ATG CTT AAA	1392
Lys Leu Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys	
450 455 460	

TCA GAA GAA GAA CGA TTA TCC ATT CAA AAT TTT AGC AAA CTT CTG AAT	1440
Ser Glu Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn	
465 470 475	
GAC AAC ATT TTT CAT ATG TCT TTA TTG GCG TGC GCT CTT GAG GTT GTA	1488
Asp Asn Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val	
480 485 490	
ATG GCC ACA TAT AGC AGA AGT ACA TCT CAG AAT CTT GAT TCT GGA ACA	1536
Met Ala Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr	
495 500 505 510	
GAT TTG TCT TTC CCA TGG ATT CTG AAT GTG CTT AAT TTA AAA GCC TTT	1584
Asp Leu Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe	
515 520 525	
GAT TTT TAC AAA GTG ATC GAA AGT TTT ATC AAA GCA GAA GGC AAC TTG	1632
Asp Phe Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu	
530 535 540	
ACA AGA GAA ATG ATA AAA CAT TTA GAA CGA TGT GAA CAT CGA ATC ATG	1680
Thr Arg Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met	
545 550 555	
GAA TCC CTT GCA TGG CTC TCA GAT TCA CCT TTA TTT GAT CTT ATT AAA	1728
Glu Ser Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys	
560 565 570	
CAA TCA AAG GAC CGA GAA GGA CCA ACT GAT CAC CTT GAA TCT GCT TGT	1776
Gln Ser Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys	
575 580 585 590	
CCT CTT AAT CTT CCT CTC CAG AAT AAT CAC ACT GCA GCA GAT ATG TAT	1824
Pro Leu Asn Leu Pro Leu Gln Asn Asn His Thr Ala Ala Asp Met Tyr	
595 600 605	
CTT TCT CCT GTA AGA TCT CCA AAG AAA AAA GGT TCA ACT ACG CGT GTA	1872
Leu Ser Pro Val Arg Ser Pro Lys Lys Lys Gly Ser Thr Thr Arg Val	
610 615 620	
AAT TCT ACT GCA AAT GCA GAG ACA CAA GCA ACC TCA GCC TTC CAG ACC	1920
Asn Ser Thr Ala Asn Ala Glu Thr Gln Ala Thr Ser Ala Phe Gln Thr	
625 630 635	
CAG AAG CCA TTG AAA TCT ACC TCT CTT TCA CTG TTT TAT AAA AAA GTG	1968
Gln Lys Pro Leu Lys Ser Thr Ser Leu Ser Leu Phe Tyr Lys Lys Val	
640 645 650	
TAT CGG CTA GCC TAT CTC CGG CTA AAT ACA CTT TGT GAA CGC CTT CTG	2016
Tyr Arg Leu Ala Tyr Leu Arg Leu Asn Thr Leu Cys Glu Arg Leu Leu	
655 660 665 670	

TCT GAG CAC CCA GAA TTA GAA CAT ATC ATC TGG ACC CTT TTC CAG CAC	2064
Ser Glu His Pro Glu Leu Glu His Ile Ile Trp Thr Leu Phe Gln His	
675 680 685	
ACC CTG CAG AAT GAG TAT GAA CTC ATG AGA GAC AGG CAT TTG GAC CAA	2112
Thr Leu Gln Asn Glu Tyr Glu Leu Met Arg Asp Arg His Leu Asp Gln	
690 695 700	
ATT ATG ATG TGT TCC ATG TAT GGC ATA TGC AAA GTG AAG AAT ATA GAC	2160
Ile Met Met Cys Ser Met Tyr Gly Ile Cys Lys Val Lys Asn Ile Asp	
705 710 715	
CTT AAA TTC AAA ATC ATT GTA ACA GCA TAC AAG GAT CTT CCT CAT GCT	2208
Leu Lys Phe Lys Ile Ile Val Thr Ala Tyr Lys Asp Leu Pro His Ala	
720 725 730	
GTT CAG GAG ACA TTC AAA CGT GTT TTG ATC AAA GAA GAG GAG TAT GAT	2256
Val Gln Glu Thr Phe Lys Arg Val Leu Ile Lys Glu Glu Glu Tyr Asp	
735 740 745 750	
TCT ATT ATA GTA TTC TAT AAC TCG GTC TTC ATG CAG AGA CTG AAA ACA	2304
Ser Ile Ile Val Phe Tyr Asn Ser Val Phe Met Gln Arg Leu Lys Thr	
755 760 765	
AAT ATT TTG CAG TAT GCT TCC ACC AGG CCC CCT ACC TTG TCA CCA ATA	2352
Asn Ile Leu Gln Tyr Ala Ser Thr Arg Pro Pro Thr Leu Ser Pro Ile	
770 775 780	
CCT CAC ATT CCT CGA AGC CCT TAC AAG TTT CCT AGT TCA CCC TTA CGG	2400
Pro His Ile Pro Arg Ser Pro Tyr Lys Phe Pro Ser Ser Pro Leu Arg	
785 790 795	
ATT CCT GGA GGG AAC ATC TAT ATT TCA CCC CTG AAG AGT CCA TAT AAA	2448
Ile Pro Gly Gly Asn Ile Tyr Ile Ser Pro Leu Lys Ser Pro Tyr Lys	
800 805 810	
ATT TCA GAA GGT CTG CCA ACA CCA ACA AAA ATG ACT CCA AGA TCA AGA	2496
Ile Ser Glu Gly Leu Pro Thr Pro Thr Lys Met Thr Pro Arg Ser Arg	
815 820 825 830	
ATC TTA GTA TCA ATT GGT GAA TCA TTC GGG ACT TCT GAG AAG TTC CAG	2544
Ile Leu Val Ser Ile Gly Glu Ser Phe Gly Thr Ser Glu Lys Phe Gln	
835 840 845	
AAA ATA AAT CAG ATG GTA TGT AAC AGC GAC CGT GTG CTC AAA AGA AGT	2592
Lys Ile Asn Gln Met Val Cys Asn Ser Asp Arg Val Leu Lys Arg Ser	
850 855 860	
GCT GAA GGA AGC AAC CCT CCT AAA CCA CTG AAA AAA CTA CGC TTT GAT	2640
Ala Glu Gly Ser Asn Pro Pro Lys Pro Leu Lys Lys Leu Arg Phe Asp	
865 870 875	

ATT GAA GGA TCA GAT GAA GCA GAT GGA AGT AAA CAT CTC CCA GGA GAG	2688
Ile Glu Gly Ser Asp Glu Ala Asp Gly Ser Lys His Leu Pro Gly Glu	
880 885 890	
TCC AAA TTT CAG CAG AAA CTG GCA GAA ATG ACT TCT ACT CGA ACA CGA	2736
Ser Lys Phe Gln Gln Lys Leu Ala Glu Met Thr Ser Thr Arg Thr Arg	
895 900 905 910	
ATG CAA AAG CAG AAA ATG AAT GAT AGC ATG GAT ACC TCA AAC AAG GAA	2784
Met Gln Lys Gln Lys Met Asn Asp Ser Met Asp Thr Ser Asn Lys Glu	
915 920 925	
GAG AAA TGAGGATCTC AGGACCTTGG TGGACACTGT GTACACCTCT GGATTCATTG	2840
Glu Lys	
TCTCTCACAG ATGTGACTGT ATAACTTTCC CAGGTTCTGT TTATGGCCAC ATTTAATATC	2900
TTCAGCTCTT TTTGTGGATA TAAAATGTGC AGATGCAATT GTTTGGGTGA TTCCTAAGCC	2960
ACTTGAAATG TTAGTCATTG TTATTTATAC AAGATTGAAA ATCTTGTGTA AATCCTGCCA	3020
TTTAAAAAGT TGTAGCAGAT TGTTTCCTCT TCCAAAGTAA AATTGCTGTG CTTTATGGAT	3080
AGTAAGAATG GCCCTAGAGT GGGAGTCCTG ATAACCCAGG CCTGTCTGAC TACTTTGCCT	3140
TCTTTTGTAG CATATAGGTG ATGTTTGCTC TTGTTTTTAT TAATTTATAT GTATATTTTT	3200
TTAATTTAAC ATGAACACCC TTAGAAAATG TGTCCTATCT ATCTTCCAAA TGCAATTTGA	3260
TTGACTGCCC ATTCACCAAA ATTATCCTGA ACTCTTCTGC AAAAAATGGAT ATTATTAGAA	3320
ATTAGAAAAA AATTACTAAT TTTACACATT AGATTTTATT TTACTATTGG AATCTGATAT	3380
ACTGTGTGCT TGTTTTATAA AATTTTGCTT TTAATTAAAT AAAAGCTGGA AGCAAAGTAT	3440
AACCATATGA TACTATCATA CTA CTACTGAAAC AGATTTTCATA CCTCAGAATG TAAAAGAACT	3500
TACTGATTAT TTTCTTCATC CAACTTATGT TTTTAAATGA GGATTATTGA TAGT	3554

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 928 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Met Pro Pro Lys Thr Pro Arg Lys Thr Ala Ala Thr Ala Ala Ala Ala
1 5 10 15

Ala	Ala	Glu	Pro	Pro	Ala	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Glu	Glu	Asp	20	25	30
Pro	Glu	Gln	Asp	Ser	Gly	Pro	Glu	Asp	Leu	Pro	Leu	Val	Arg	Leu	Glu	35	40	45
Phe	Glu	Glu	Thr	Glu	Glu	Pro	Asp	Phe	Thr	Ala	Leu	Cys	Gln	Lys	Leu	50	55	60
Lys	Ile	Pro	Asp	His	Val	Arg	Glu	Arg	Ala	Trp	Leu	Thr	Trp	Glu	Lys	65	70	75
Val	Ser	Ser	Val	Asp	Gly	Val	Leu	Gly	Gly	Tyr	Ile	Gln	Lys	Lys	Lys	85	90	95
Glu	Leu	Trp	Gly	Ile	Cys	Ile	Phe	Ile	Ala	Ala	Val	Asp	Leu	Gly	Asp	100	105	110
Met	Ser	Phe	Thr	Phe	Thr	Glu	Leu	Gln	Lys	Asn	Ile	Glu	Ile	Ser	Val	115	120	125
His	Lys	Phe	Phe	Asn	Leu	Leu	Lys	Glu	Ile	Asp	Thr	Ser	Thr	Lys	Val	130	135	140
Asp	Asn	Ala	Met	Ser	Arg	Leu	Leu	Lys	Lys	Tyr	Asp	Val	Leu	Phe	Ala	145	150	155
Leu	Phe	Ser	Lys	Leu	Glu	Arg	Thr	Cys	Glu	Leu	Ile	Tyr	Leu	Thr	Gln	165	170	175
Pro	Ser	Ser	Ser	Ile	Ser	Thr	Glu	Ile	Asn	Ser	Ala	Leu	Val	Leu	Lys	180	185	190
Val	Ser	Trp	Ile	Thr	Phe	Leu	Leu	Ala	Lys	Gly	Glu	Val	Leu	Gln	Met	195	200	205
Glu	Asp	Asp	Leu	Val	Ile	Ser	Phe	Gln	Leu	Met	Leu	Cys	Val	Leu	Asp	210	215	220
Tyr	Phe	Ile	Lys	Leu	Ser	Pro	Pro	Met	Leu	Leu	Lys	Glu	Pro	Tyr	Lys	225	230	235
Thr	Ala	Val	Ile	Pro	Ile	Asn	Gly	Ser	Pro	Arg	Thr	Pro	Arg	Arg	Gly	245	250	255
Gln	Asn	Arg	Ser	Ala	Arg	Ile	Ala	Lys	Gln	Leu	Glu	Asn	Asp	Thr	Arg	260	265	270
Ile	Ile	Glu	Val	Leu	Cys	Lys	Glu	His	Glu	Cys	Asn	Ile	Asp	Glu	Val	275	280	285
Lys	Asn	Val	Tyr	Phe	Lys	Asn	Phe	Ile	Pro	Phe	Met	Asn	Ser	Leu	Gly	290	295	300

Leu Val Thr Ser Asn Gly Leu Pro Glu Val Glu Asn Leu Ser Lys Arg
 305 310 315 320
 Tyr Glu Glu Ile Tyr Leu Lys Asn Lys Asp Leu Asp Ala Arg Leu Phe
 325 330 335
 Leu Asp His Asp Lys Thr Leu Gln Thr Asp Ser Ile Asp Ser Phe Glu
 340 345 350
 Thr Gln Arg Thr Pro Arg Lys Ser Asn Leu Asp Glu Glu Val Asn Val
 355 360 365
 Ile Pro Pro His Thr Pro Val Arg Thr Val Met Asn Thr Ile Gln Gln
 370 375 380
 Leu Met Met Ile Leu Asn Ser Ala Ser Asp Gln Pro Ser Glu Asn Leu
 385 390 395 400
 Ile Ser Tyr Phe Asn Asn Cys Thr Val Asn Pro Lys Glu Ser Ile Leu
 405 410 415
 Lys Arg Val Lys Asp Ile Gly Tyr Ile Phe Lys Glu Lys Phe Ala Lys
 420 425 430
 Ala Val Gly Gln Gly Cys Val Glu Ile Gly Ser Gln Arg Tyr Lys Leu
 435 440 445
 Gly Val Arg Leu Tyr Tyr Arg Val Met Glu Ser Met Leu Lys Ser Glu
 450 455 460
 Glu Glu Arg Leu Ser Ile Gln Asn Phe Ser Lys Leu Leu Asn Asp Asn
 465 470 475 480
 Ile Phe His Met Ser Leu Leu Ala Cys Ala Leu Glu Val Val Met Ala
 485 490 495
 Thr Tyr Ser Arg Ser Thr Ser Gln Asn Leu Asp Ser Gly Thr Asp Leu
 500 505 510
 Ser Phe Pro Trp Ile Leu Asn Val Leu Asn Leu Lys Ala Phe Asp Phe
 515 520 525
 Tyr Lys Val Ile Glu Ser Phe Ile Lys Ala Glu Gly Asn Leu Thr Arg
 530 535 540
 Glu Met Ile Lys His Leu Glu Arg Cys Glu His Arg Ile Met Glu Ser
 545 550 555 560
 Leu Ala Trp Leu Ser Asp Ser Pro Leu Phe Asp Leu Ile Lys Gln Ser
 565 570 575
 Lys Asp Arg Glu Gly Pro Thr Asp His Leu Glu Ser Ala Cys Pro Leu
 580 585 590

305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590

Asn	Leu	Pro	Leu	Gln	Asn	Asn	His	Thr	Ala	Ala	Asp	Met	Tyr	Leu	Ser	595	600	605	
Pro	Val	Arg	Ser	Pro	Lys	Lys	Lys	Gly	Ser	Thr	Thr	Arg	Val	Asn	Ser	610	615	620	
Thr	Ala	Asn	Ala	Glu	Thr	Gln	Ala	Thr	Ser	Ala	Phe	Gln	Thr	Gln	Lys	625	630	635	640
Pro	Leu	Lys	Ser	Thr	Ser	Leu	Ser	Leu	Phe	Tyr	Lys	Lys	Val	Tyr	Arg	645	650	655	
Leu	Ala	Tyr	Leu	Arg	Leu	Asn	Thr	Leu	Cys	Glu	Arg	Leu	Leu	Ser	Glu	660	665	670	
His	Pro	Glu	Leu	Glu	His	Ile	Ile	Trp	Thr	Leu	Phe	Gln	His	Thr	Leu	675	680	685	
Gln	Asn	Glu	Tyr	Glu	Leu	Met	Arg	Asp	Arg	His	Leu	Asp	Gln	Ile	Met	690	695	700	
Met	Cys	Ser	Met	Tyr	Gly	Ile	Cys	Lys	Val	Lys	Asn	Ile	Asp	Leu	Lys	705	710	715	720
Phe	Lys	Ile	Ile	Val	Thr	Ala	Tyr	Lys	Asp	Leu	Pro	His	Ala	Val	Gln	725	730	735	
Glu	Thr	Phe	Lys	Arg	Val	Leu	Ile	Lys	Glu	Glu	Glu	Tyr	Asp	Ser	Ile	740	745	750	
Ile	Val	Phe	Tyr	Asn	Ser	Val	Phe	Met	Gln	Arg	Leu	Lys	Thr	Asn	Ile	755	760	765	
Leu	Gln	Tyr	Ala	Ser	Thr	Arg	Pro	Pro	Thr	Leu	Ser	Pro	Ile	Pro	His	770	775	780	
Ile	Pro	Arg	Ser	Pro	Tyr	Lys	Phe	Pro	Ser	Ser	Pro	Leu	Arg	Ile	Pro	785	790	795	800
Gly	Gly	Asn	Ile	Tyr	Ile	Ser	Pro	Leu	Lys	Ser	Pro	Tyr	Lys	Ile	Ser	805	810	815	
Glu	Gly	Leu	Pro	Thr	Pro	Thr	Lys	Met	Thr	Pro	Arg	Ser	Arg	Ile	Leu	820	825	830	
Val	Ser	Ile	Gly	Glu	Ser	Phe	Gly	Thr	Ser	Glu	Lys	Phe	Gln	Lys	Ile	835	840	845	
Asn	Gln	Met	Val	Cys	Asn	Ser	Asp	Arg	Val	Leu	Lys	Arg	Ser	Ala	Glu	850	855	860	
Gly	Ser	Asn	Pro	Pro	Lys	Pro	Leu	Lys	Lys	Leu	Arg	Phe	Asp	Ile	Glu	865	870	875	880

